Entwicklung und Implementierung PC-gestützter akustischer Analyseverfahren für die klinische Diagnostik neurogener Sprechstörungen

Michael Merk

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	2000	Dissertation

4. TITLE AND SUBTITLE

Entwicklung und Implementierung PC-gestuetzter akustischer Analyseverfahren fuer die klinische Diagnostik neurogener Sprechstoerungen

(Development and Implementation of PC-based Acoustic Analytic Procedures for the Clinical Diagnosis of Neurogenic Speech Defects)

5. FUNDING NUMBERS

6. AUTHOR(S)

Michael Merk

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Universitaet fuer der Bundeswehr Muenchen
Fakultaet fuer Elektrotechnik (Energie- und Informationstechnik)

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES

UNIBW

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Text in German, 282 pages.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Category A; Public Release

12b. DISTRIBUTION CODE

ABSTRACT (Maximum 200 words)

Approximately two percent of the population suffers from brain damage. Some of these cases lead to speech defects. The diagnosis of neurogenic speech defects is a speciality of clinical neuropsychology. The typical approach is subjective. Technical acoustic evaluation procedures are rarely used in diagnosis and therapy. This dissertation describes a "modular diagnostic system for speech defects" that supports the standard diagnosis of neurogenic speech defects based on the acoustic analysis of the speech signal and the compilation of diagnostic parameters.

Machine assisted translation.

14. SUBJECT TERMS

UNIBW, Germany, Neurogenic speech defects, Brain damage, Speech signal, Acoustic evaluation procedures

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

UNIVERSITÄT DER BUNDESWEHR MÜNCHEN
Fakultät für Elektrotechnik
(Energie- und Informationstechnik)

Entwicklung und Implementierung PC-gestützter akustischer Analyseverfahren für die klinische Diagnostik neurogener Sprechstörungen

Michael Merk

Vorsitzender des Promotionsausschusses:	Prof. Dr.-Ing. K. Landes
1. Berichterstatter:	Prof. Dr.-Ing. K. Tröndle
2. Berichterstatter:	Prof. Dr.-Ing. W. Wolf
3. Berichterstatter:	Dr.rer.nat. W. Ziegler

Tag der Prüfung 30.03.2000

Mit der Promotion erlangter akademischer Grad:

Doktor-Ingenieur
(Dr.-Ing.)

Neubiberg, im April 2000

Gewidmet meiner lieben Martina ...

Vorwort

Die vorliegende Arbeit entstand während meiner nebenberuflichen Tätigkeit als externer wissenschaftlicher Mitarbeiter am Institut für Nachrichtentechnik der Universität der Bundeswehr München in Kooperation mit der Entwicklungsgruppe klinische Neuropsychologie des Städtischen Krankenhauses München-Bogenhausen.

An dieser Stelle möchte ich mich herzlich bedanken bei meinem Doktorvater Herrn Prof. Dr.-Ing. Kurt Tröndle, der mir 1994 diese Arbeit ermöglichte. Weiterhin bedanke ich mich bei Herrn Prof. Dr.-Ing. Klaus Landes für die Übernahme des Vorsitzes und bei Herrn Prof. Dr.-Ing. Werner Wolf für die Übernahme des Koreferates.
Herrn Dr. rer.nat. Wolfram Ziegler danke ich für nun 6 Jahre einer sehr kollegialen und freundschaftlichen Kooperation. Sein Bestreben, ohne jeden Eigennutz die neuropsychologische Forschung und Entwicklung im Sinne einer verbesserten Patientenversorgung möglichst effektiv voranzutreiben, war mir stets ein Vorbild. Für die Betreuung meiner Arbeit zu jeder Zeit und an jedem Ort danke ich ihm besonders.
Danken möchte ich auch allen Mitarbeitern des Instituts, ganz besonders aber Frau Frendl und Herrn Besthorn für viele Jahre kontinuierlicher Unterstützung in materieller, organisatorischer und persönlicher Hinsicht. Die beiden haben mich immer so einbezogen, als wäre ich ein fester Mitarbeiter vor Ort gewesen.

Meinem lieben Schwiegervater Herrn Karl Baier danke ich sehr herzlich für das zeitaufwendige Korrekturlesen dieser Arbeit. Sollte jetzt dennoch der eine oder andere Tippfehler enthalten sein, dann trage ich dafür allein die Verantwortung. Aus meiner Familie danke ich auch Herrn Harald Mix für seine kräftige Unterstützung bei der Übersetzung meiner Zusammenfassung in die englische Sprache.

Und nun zu meiner geliebten Ehefrau Martina: Ich habe für diese Arbeit fast meine gesamte Freizeit der letzten 6 Jahre geopfert und damit auch gleichzeitig ihr vorenthalten. Trotzdem hat sie mich stets auf meinem Weg unterstützt und begleitet. Ich werde ihr immer dafür dankbar sein.

Mauerstetten, im April 2000

Michael Mohr

Inhaltsverzeichnis

Einführung

Ungefähr 2 Prozent der Bevölkerung der westlichen Welt leiden an den Nachwirkungen einer Hirnschädigung. Sie kann verursacht sein durch Schlaganfälle, Hirnblutungen, Schädel-Hirn-Verletzungen, traumatische Erkrankungen, Hirntumore oder entzündliche Prozesse im Gehirn. In vielen Fällen sind mit einer solchen Schädigung **neurogene Sprechstörungen** verbunden, die alle funktionellen Strukturen des Sprechvorganges wie Sprechatmung, Stimme und Artikulation betreffen können. Durch die Einschränkung der kommunikativen Fähigkeiten wird eine solche Behinderung durch den Patienten im allgemeinen als besonders schwerwiegend empfunden.

Die **Diagnostik** neurogener Sprechstörungen ist ein Aufgabenbereich der **klinischen Neuropsychologie**. Hierbei kommen heute noch vorwiegend auditive Verfahren zum Einsatz, d.h., ein geschulter Hörer beurteilt einen Patienten nach seinem subjektiven Höreindruck. Akustisch-phonetische Verfahren, also der unterstützende Einsatz technischer Systeme, tragen hier zu einer Objektivierung des klinischen Bildes bei. Obwohl die Verwendung akustischer Parameter in der Beschreibung neurogener Sprechstörungen mittlerweile zum Standard wissenschaftlicher Untersuchungen gehört, ist die tatsächliche Bedeutung dieser Parameter in der klinischen Diagnostik, gemessen an diesen Entwicklungen, immer noch gering. Nur **in wenigen klinischen Einrichtungen** werden neurogene Sprechstörungen routinemäßig akustisch evaluiert, und vor allem in der Quantifizierung von Therapieeffekten fehlt es an Belegen durch valide akustische Parameter. Ein Grund für diesen Mangel kann darin gesehen werden, daß die meisten Verfahren für die Routineanwendung immer noch **zu komplex** sind, und die Aussagekraft vieler Parameter **nicht ausreichend transparent** ist.

Mit dem im Rahmen dieser Arbeit entstandenen **Modularen Diagnostik-System**

für Sprechstörungen (MODIAS) soll dem Diagnostiker erstmals ein **umfassendes Instrument zur Standarddiagnostik** neurogener Sprechstörungen anhand akustischer Analysen des Sprachsignals und **zur Befunderstellung anhand akustischer Parameter** an die Hand gegeben werden. Der Anspruch der klinischen Verwendbarkeit bedingte dabei unter anderem die Integration aller relevanten Aspekte einer klinischen Standarddiagnostik von der Präsentation des Stimulusmaterials über die Sprachaufzeichnung mit anschließender Berechnung akustischer Parameter bis hin zur Ergebnispräsentation. Es sollten nur klinisch relevante Parameter geliefert werden, die auch geeignet sind, valide Aussagen über die wichtigsten Aspekte wie Artikulation, Stimme und Prosodie zu treffen. Sämtliche akustischen Parameter sollten technisch bezüglich Stabilität und Exaktheit optimiert sein, auch unter bewußter Inkaufnahme höherer Rechenzeiten. Sämtliche Berechnungen sollten dabei eine nur minimale Benutzerinteraktion erfordern sowie klare und leicht interpretierbare Ergebnisse liefern. Zusätzlich sollten **neueste Methoden der Digitalen Signalverarbeitung** Anwendung finden.

Die Lösung einer solchen **interdisziplinären Aufgabenstellung** kann nur durch ein Zusammenwirken neuropsychologischer Expertise mit ingenieurwissenschaftlicher Methodik bewältigt werden. Durch die Vielzahl involvierter Fachdisziplinen wie der Neuropsychologie, Neurolinguistik und Phonetik sowie der Elektrotechnik und der Informatik ist hierbei ein ständiger Dialog der verschiedenen Experten erforderlich, besonders in den iterativen Phasen der klinischen Erprobungen.
Die Arbeitsgruppe *Sprach- und Sprechstörungen* der **Entwicklungsgruppe klinische Neuropsychologie des Städtischen Krankenhauses München-Bogenhausen (EKN)** beschäftigt sich mit Fragen der Diagnostik und der Behandlung von sprechmotorischen Störungen bei dysarthrischen und aphasischen Patienten, aphasischen Störungen vor allem der phonologischen Verarbeitung und des Wortabrufs und nicht-aphasischen, primär kognitiv bedingten Störungen der Textverarbeitung bei Patienten mit Hirnschädigung. Zu den Hauptaspekten dieser Aktivitäten zählt die Entwicklung und Evaluation von klinisch einsetzbaren computergesteuerten Verfahren für die Anwendung in der klinischen Versorgung.
Die wissenschaftliche Arbeitsgruppe *Computersysteme für medizinische Diagnose, Therapie und Lebenshilfe* des **Instituts für Nachrichtentechnik der Universität der Bundeswehr München** leistet seit vielen Jahren Beiträge zur Hilfestellung

behinderter Menschen und zur Unterstützung von Medizinern in ihrer täglichen Arbeit. Durch langjährige und fruchtbare Kooperation dieser beiden Institutionen waren bereits der *Testplatz zur Untersuchung von Sprechstörungen (TUS)* und das System *PhonX* entstanden. Der TUS wird derzeit überwiegend zur Durchführung von Verständlichkeitsuntersuchungen nach dem Münchner Verständlichkeitsprofil (MVP) eingesetzt und bildet die technische Grundlage einer Studie zur Sprechapraxietherapie. PhonX wird überwiegend zur Steuerung von neurophonetischen Experimenten zur Sprachproduktion und auditiven Sprachverarbeitung verwendet. Das hier vorgestellte Diagnostik-System MODIAS ist nun das Ergebnis einer weiteren **über fünfjährigen** Kooperation.

Die vorliegende **schriftliche Arbeit** ist so aufgebaut, daß ein klinischer Diagnostiker schnell und zuverlässig das Potential des MODIAS-Systems erfassen und auf seine konkreten Bedürfnisse projizieren kann. Er sollte anhand dieser Arbeit eine vollständige Patientenuntersuchung durchführen und deren Ergebnisse uneingeschränkt interpretieren können. Auf der anderen Seite sollte dem Ingenieur ein hinreichender Einblick in die technischen Verfahren gewährt werden, und zwar in einem Detaillierungsgrad, der ihn bei vergleichbaren Projekten zu ähnlichen oder weiter optimierten Lösungsansätzen inspirieren kann.

Kapitel 1 schafft zunächst die anatomischen und physiologischen Grundlagen für das Verständnis des Sprechvorgangs. Danach folgen die Grundzüge der Phonetik, die sich mit der Beschreibung und der Klassifikation von Sprachlauten beschäftigt. Anschließend werden die akustischen Grundlagen aufgezeigt, also die Eigenschaften des Sprachschalls in Abhängigkeit von der übertragenen Information. Abschließend werden die suprasegmentalen bzw. prosodischen Merkmale der Sprache näher beleuchtet.

Kapitel 2 beschäftigt sich mit dem Themengebiet neurogener Sprechstörungen. Nach einer neuropsychologischen Definition werden zunächst die Ätiologien und die verschiedenen Syndrome vorgestellt. Anschließend werden die aktuellen Verfahren in der Sprechstörungsdiagnostik aufgezeigt.

Kapitel 3 widmet sich der Beschreibung des Diagnostischen und Technischen Konzepts des MODIAS-Systems. Hier soll ein Überblick geschaffen werden, welche Probleme in der klinischen Sprechstörungsdiagnostik vorherrschen und welche Überlegungen bei der

Konzeption des MODIAS-Systems Berücksichtigung fanden.

Die Realisierung des MODIAS-Systems wird in den **Kapiteln 4-7** getrennt nach den 4 enthaltenen Einzelmodulen vorgestellt. Dabei ist jedes dieser Kapitel identisch strukturiert:
Ausgehend von der jeweils beleuchteten diagnostischen Fragestellung wird zunächst der Aufbau und die Gestaltung des Programm-Moduls vorgestellt, also im wesentlichen der Ablauf und die Durchführung einer Untersuchung und die Interpretation der Ergebnisse **ohne technische Details**. Damit soll ein **Diagnostiker** in die Lage versetzt werden, das jeweilige Modul unmittelbar in seinem klinischen Umfeld einzusetzen.
Danach folgt der **technisch-wissenschaftliche Anteil** bzw. die Vorstellung sämtlicher technischer Verfahren und Algorithmen, die für den Anwender nicht sichtbar ablaufen und die Funktionalität des Systems gewährleisten. Die Darstellung erfolgt dabei sehr transparent und in einem hinreichend hohen Detaillierungsgrad, um einem **Ingenieur** bei vergleichbaren Projekten konkrete Hilfe anzubieten.
Am Ende des Kapitels werden jeweils einige **klinische Anwendungen** präsentiert, die sowohl dem Diagnostiker als auch dem Ingenieur die Einsatzmöglichkeiten und die bisherigen Resultate aufzeigen und damit einen Eindruck von der **breit gefächerten klinischen Nutzung** vermitteln sollen.

1. Der Sprechvorgang

Der Vorgang des *Sprechens* entsteht aus dem komplexen Zusammenwirken der Komponenten *Atmung*, *Phonation* und *Artikulation* ([Bas86], [Lie77], [Per86], [Vog87], [Vog88]). Die *Ausatmung* muß dabei kontinuierlich erfolgen; sie ist gegenüber der *Ruheatmung* um ca. 1/3 verlängert. Dies wird durch eine Beschleunigung des *Einatmens* beim Sprechen erreicht, wobei die Gesamtdauer des Atemzyklus beim Sprechen und in der Ruhe etwa gleich ist und im Durchschnitt ca. 5 Sekunden beträgt ([Poe89]).

Der Luftstrom wird im Stimmapparat moduliert, d.h., in Turbulenzen oder periodische Schwingungen versetzt (*Phonation*).

Beim Passieren des Rachen- und Mundraumes kommt es je nach Öffnungsgrad und Zungenposition zu Turbulenzen und plötzlichen Unterbrechungen des Luftstroms. Erst durch *Resonanzereignisse* ergeben sich hierbei die für die Sprachlaute charakteristischen Klang- und Geräuschmuster. Die Veränderungen der Klang- bzw. Geräuscheigenschaften werden durch spezifische Bewegungen der *Artikulatoren* erreicht.
Der Sprechvorgang erfordert dabei die genaue zeitliche Abstimmung aller am Sprechvorgang beteiligter Muskelgruppen, das sind

- Zwerchfell, Bauch- und Brustmuskulatur,
- Kehlkopfmuskulatur,
- Zungenbein- und Rachenmuskulatur,
- Zungenmuskulatur,
- Mund- und Kiefermuskulatur.

Für normales Sprechen gilt, daß Art und Ausmaß der Respirations-, Phonations- und Artikulationsvorgänge exakt koordiniert und die artikulatorischen Ziele im Mundraum rasch erreicht werden müssen. Diese (zeitliche) Koordination variiert dabei nach *Sprechgeschwindigkeit*, *Sprechrhythmus* und *Sprechmelodie*.

1.1. Anatomische und physiologische Grundlagen

Die *Sprechwerkzeuge* sind diejenigen Organe, die am Sprechvorgang beteiligt sind. Die Abbildung 1.1 zeigt im Überblick, daß hierzu nicht nur Mund und Hals, sondern auch Nase, Luftröhre und Lunge zu zählen sind.

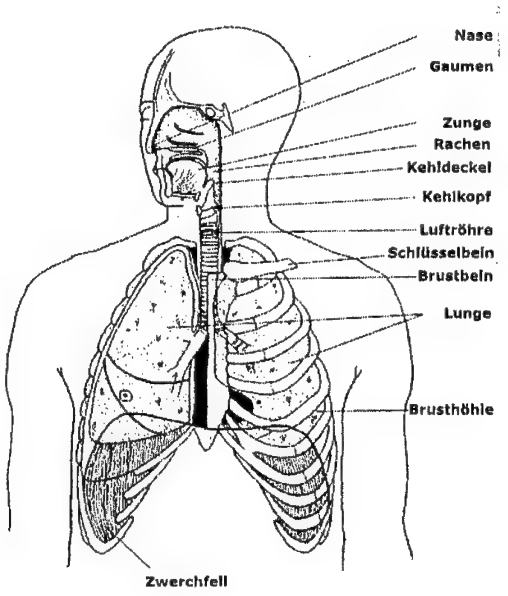

Abbildung 1.1.: Die Sprechwerkzeuge des Menschen im Überblick ([Cry95]).

1.1.1. Atemorgane

Laute sind Schallwellen, zu deren Erzeugung grundsätzlich eine Energiequelle erforderlich ist. Beim Sprechvorgang übernehmen die *Atemorgane* diese Aufgabe, indem sie die vor dem Sprechvorgang eingeatmete Luft als *pulmonale* Luft kontinuierlich und gegenüber normaler *Exspiration* (Ausatmung) zeitlich verlängert wieder ausströmen lassen und so den alveolaren Druck aufbauen, der für die anschließende Phonation benötigt wird.
Unter den Atemorganen versteht man das *Thoraxskelett*, die *Atemmuskulatur* und die *Lunge*.

1.1.2. Stimmapparat und Stimmbildung

Zum Sprechen wird pulmonale Luft durch verschiedene Organe des Vokaltrakts in Schwingungen versetzt. Für Sprachlaute ist die wichtigste Schwingungsquelle der Kehlkopf (Larynx) am oberen Teil der Luftröhre (siehe Abbildung 1.4).
Der Kehlkopf ist ein röhrenförmiges Knorpelgerüst, das durch Bänder und Membranen beweglich zusammengehalten wird. Er läßt sich vorne am Hals leicht durch den vorstehenden *Schildknorpel* ertasten. Zum Kehlkopf gehören außerdem der *Ringknorpel* und die beiden *Stell-* oder *Gießbeckenknorpel*. Die Bewegungen der Knorpel steuern die Schwingungen der *Stimmbänder*.
Als Stimmbänder bezeichnet man den oberen freien Rand der Stimmfalten, die sich von der Rückfläche des Schildknorpels zum Stellknorpel ziehen und jeweils das eigentliche Stimmband und den Stimm-Muskel enthalten (Abbildung 1.2).
Die Öffnung zwischen den Stimmbändern wird als *Stimmritze* (Glottis) bezeichnet.
Die sehr komplexen Wechselwirkungen zwischen den Kehlkopfmuskeln verändern die Spannung, Elastizität, Höhe, Breite, Länge und Dicke der Stimmbänder. Beim Sprechen vollziehen sich diese Anpassungen mit hoher Geschwindigkeit, und sie haben unterschiedliche Auswirkungen auf die akustischen Eigenschaften der erzeugten Laute.
Der wichtigste Effekt ist hierbei die Erzeugung hörbarer Schwingungen. Ein stimmhafter Laut entsteht, wenn die Stimmritze zu einem Spalt verengt ist und die Stimmbänder schwingen. Alle Vokale und die meisten Konsonanten nutzen dieses Prinzip. Jedes Öffnen bzw. Schließen der Stimmritze erzeugt dabei einen Schwingungsimpuls.
Die Frequenz der Stimmbandschwingungen kann dabei innerhalb bestimmter Grenzen frei

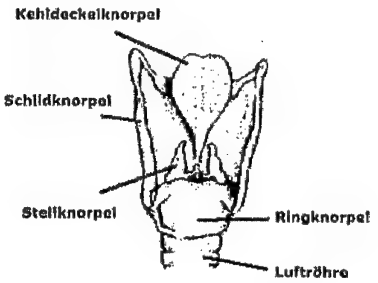

Abbildung 1.2.: Aufbau des Kehlkopfs in der Rückansicht ([Cry95]).

geändert, und damit können Tonhöhe und Lautstärke variiert werden. Bei Männern beträgt die mittlere Stimmfrequenz (auch: *Grundfrequenz* bzw. *Stimmbandgrundfrequenz*) ca. 120 Hz, bei Frauen ca. 220 Hz und liegt damit knapp eine Oktave höher.
Abbildung 1.3 zeigt die Stimmbandbewegung von einem Glottisverschluß zum nächsten bei einem stimmhaften Laut. Dieser Vorgang wiederholt sich zwischen 50 und 300 mal pro Sekunde je nach Stimmfrequenz.
Die heutige Erklärung für das Zustandekommen der Stimmbandbewegung und somit den Effekt der Stimmbildung geht auf die *myoelastisch-aerodynamische Theorie* zurück. Demnach bewirkt der Luftdruck bei jeder Einzelschwingung eine Öffnung der Stimmbänder, die anschließend sofort wieder aneinandergesaugt werden. Diese somit nicht neuronal innervierte (wie von HUSSON ursprünglich in [Hus62] postuliert), sondern aerodynamisch bedingte Verschlußbildung beruht nach der genannten Theorie sowohl auf der Eigenelastizität der Stimmlippen als auch auf dem plötzlichen Druckabfall in der Glottis (*Bernoulli-Effekt*).
VAN DEN BERG veröffentlichte hierzu im Jahre 1957 ein Grundlagenwerk ([Van57]), das diese Mechanismen innerhalb der Larynx analysiert; es bildet auch heute die Basis für das Verständnis der Stimmbildung (siehe auch [Fla58]).

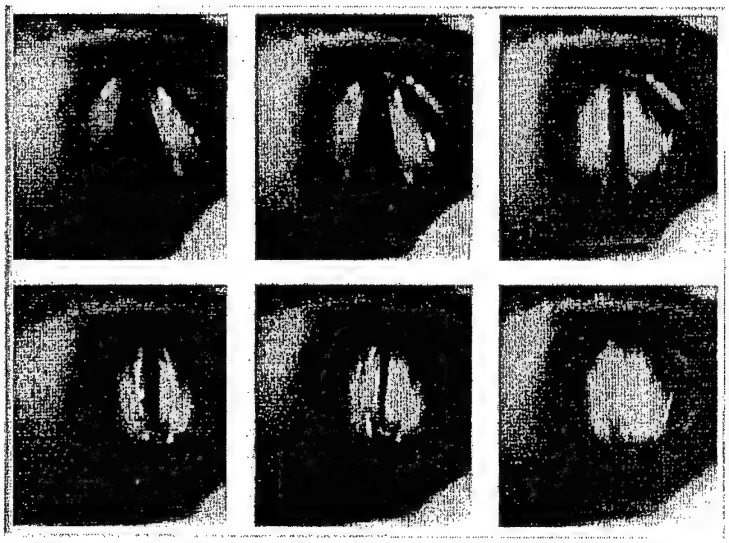

Abbildung 1.3.: Hochgeschwindigkeits-Aufnahme der Stimmbandbewegung. Die obere Reihe zeigt, wie sich die Stimmritze vor der Phonation verengt. Darunter sind die Stimmbänder während der Phonation zu erkennen: Die noch einen Spalt geöffnete Stimmritze schließt sich völlig, wodurch jeweils ein Schwingungsimpuls erzeugt wird ([Cry95]).

1.1.3. Vokaltrakt und Artikulation

Nach Passieren des Kehlkopfs bzw. der Stimmbänder gelangt der Luftstrom in ein langes, röhrenförmiges Gebilde, das sogenannte *Ansatzrohr* bzw. den *Vokaltrakt*. Die verschiedenen Luftkammern des Vokaltrakts, wie z.B. Mund-, Nasen- oder Rachenraum, beeinflussen die von den Stimmbändern erzeugten Schwingungen (es handelt sich bisher nur um ein knatterndes Geräusch) und treten mit ihnen in Resonanz, wodurch charakteristische Klang- bzw. Geräuschmuster erzeugt werden (Abbildung 1.4). Diese Resonanz ist veränderlich, weil der Vokaltrakt unterschiedliche Formen annehmen kann.
Bei der Beschreibung der Artikulation wird meist zwischen beweglichen und unbeweglichen Teilen des Vokaltrakts unterschieden, also zwischen

- *aktiven*, d.h. durch den Sprecher steuerbaren Artikulatoren, und
- *passiven Artikulatoren* bzw. *Artikulationsstellen.*

Zu den passiven Artikulatoren zählen die oberen Schneidezähne, der *Alveolarkamm* (Knochendamm hinter den oberen Zähnen) und der *harte Gaumen* (knöcherne Wölbung hinter dem Alveolarkamm).
Die aktiven Artikulatoren werden wie folgt unterschieden:

- Rachen (*Pharynx*): Muskulöse Röhre, die den Kehlkopf (*Larynx*) mit dem rückwärtigen Teil des Mund- bzw. Rachenraums verbindet. Der Rachen läßt sich weiter untergliedern in Kehlkopfrachen (*Laryngopharynx*), Mundrachen (*Oropharynx*) und Nasenrachen (*Nasopharynx*). Der Rachen kann beim Sprechvorgang verengt oder geweitet werden, was sich auf die Lautqualität auswirkt.
- Gaumensegel (*Velum*): Breites muskulöses Band im hinteren oberen Bereich des Mundraums mit dem auffälligen Merkmal eines herabhängenden Zäpfchens (*Uvula*). Das Velum kann gehoben werden (Die aspirierte Luft entweicht wegen dem *Velolaryngealverschluß* nur noch durch den Mund) oder auch abgesenkt werden (Luft entweicht durch Nase und gegebenenfalls zusätzlich durch den Mund).
- Lippen: Die Lippenbewegung wird von einer Reihe von Gesichtsmuskeln gesteuert, hauptsächlich jedoch durch den *Mundringmuskel bzw. Musculus orbicularis oris.* Sie können sich beim Sprechvorgang dicht aufeinanderlegen, in unterschiedlichem Ausmaß öffnen, spreizen, runden oder nach vorne stülpen.
- Unterkiefer: Der Unterkiefer hat in seinem Öffnungsgrad einen erheblichen Einfluß auf die Resonanzbildung und auf die Stellung der Lippen.
- Zunge: Muskelkörper, der bei der Bildung fast aller Laute beteiligt ist. Sie ist über verschiedene *extrinsische* (an benachbarten Knochen entspringende) Muskeln in 3 Hauptrichtungen bewegbar (nach vorne und oben, hinten und oben bzw. hinten und unten). Die Zungenform selbst läßt sich zusätzlich über eine Reihe *intrinsischer* (zungeninterner, nicht an Knochen sitzender) Muskeln beeinflussen.

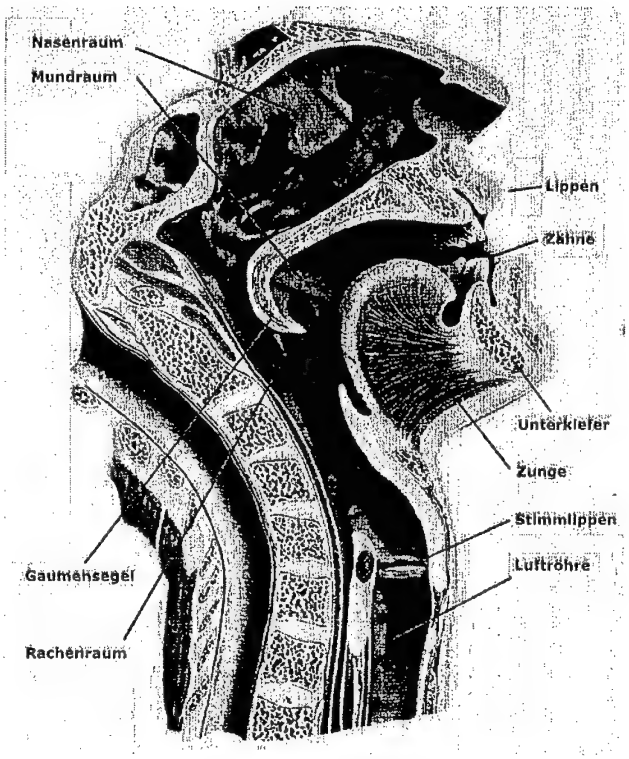

Abbildung 1.4.: Der Vokaltrakt des Menschen im Querschnitt ([Rom98]).

1.2. Phonetische Grundlagen

Die *Phonetik* befaßt sich mit der Beschreibung und Klassifikation von Sprachlauten. Diese lassen sich dabei anhand des physiologischen Mechanismus ihrer Bildung (artikulatorische Beschreibung), nach ihrer akustischen Qualität oder nach ihren auditiven Merkmalen beschreiben.
Bei der gebräuchlichsten artikulatorischen phonetischen Beschreibung werden folgende Hauptfaktoren berücksichtigt ([Cry95]):

- Luftstrom: Ursprung und Richtung des pulmonalen Luftstroms.
- Stimmbänder: *Stimmhafte* Laute entstehen, wenn die Stimmbänder schwingen; *stimmlose* Laute werden hingegen ohne Stimmbandschwingungen bei geöffneter Stimmritze gesprochen.
- Gaumensegel: Bei abgesenktem Gaumensegel strömt Luft über die Nase aus und der Laut wird als *nasal* bezeichnet. Ist es angehoben, tritt die Luft ausschließlich durch den Mund aus und der Laut ist *oral*.
- Artikulationsort: Er bezeichnet den Ort im Vokaltrakt, an dem der Verschluß bzw. die Verengung gebildet werden, die für den jeweiligen Laut bestimmend sind.
- Artikulationsart: Art der Verengung oder Bewegung an der jeweiligen Artikulationsstelle wie z.B. eine deutliche Verengung oder ein plötzlich oder allmählich aufgelöster Verschluß.
- Lippen: Stellung der Lippen; ihr kommt besonders bei Vokalen eine bedeutende Rolle zu.

Besonders wichtig ist auch die phonetische Differenzierung zwischen Konsonanten und Vokalen. Konsonanten werden hierbei als Laute definiert, die mittels eines Verschlusses im Vokaltrakt oder durch eine so starke Verengung gebildet werden, daß die Luft nur mit hörbarer Reibung entweichen kann. Vokale werden als Laute definiert, bei denen die Luft weitgehend ungehindert über Mund oder Nase ausströmen kann.

1.2.1. Phonetische Beschreibung der Vokalartikulation

Die Vokale werden üblicherweise anhand von 4 Kriterien phonetisch differenziert ([Cry95]):

- Angehobener Teil der Zunge: Unterscheidung zwischen vorderem, mittlerem oder hinterem Abschnitt.
- Grad der Zungenhebung in Richtung Gaumen: Differenzierung nach *hoher, mittlerer, tiefer* Zungenlage oder nach *geschlossener, halbgeschlossener, halboffener, offener* Zungenhebung.
- Stellung des Gaumensegels: Es ist angehoben bei oralen und abgesenkt bei nasalen Vokalen.
- Art der Lippenöffnung: Differenzierung nach dem Grad der Lippenspreizung oder Lippenrundung.

In Abbildung 1.5 ist beispielhaft die Stellung des Artikulationstraktes und die Position der Zunge für die Vokale */i/* (hohe Zungenlage im vorderen Abschnitt), */u/* (hohe Zungenlage im hinteren Abschnitt) und */a/* (tiefe Zungenlage im hinteren Abschnitt) dargestellt.

Abbildung 1.5.: Stellung des Artikulationstraktes und der Zunge bei den Vokalen */i/* (links), */u/* (mitte) und */a/* (rechts). Der Vokal */y/* (nicht dargestellt) unterscheidet sich von */i/* nur durch eine Vorstülpung der Lippen ([Fel84]).

1.2.2. Phonetische Beschreibung der Konsonantartikulation

Zu den wichtigsten Kriterien für eine phonetische Differenzierung von Konsonanten zählen ([Cry95]):

- Schwingungszustand der Stimmbänder: Schwingend (*stimmhaft*) oder nicht schwingend (*stimmlos*);
- Stellung des Gaumensegels: Angehoben (*oral*) oder abgesenkt (*nasal*);
- Artikulationsort;
- Artikulationsart.

In Deutschen werden beim Sprechen vorwiegend die folgenden Artikulationsorte (auch: Artikulationsstellen) eingesetzt (in Anlehnung an [Cry95]):

- Bilabial: Beide Lippen sind an der Artikulation beteiligt (z.B. */b/* in *Bus*).
- Labiodental: Der Laut wird mit der Unterlippe an der oberen Zahnreihe erzeugt (z.B. */f/* in *Fisch*).
- Alveolar: Der Laut wird mit der Zungenspitze an den Alveolaren artikuliert (z.B. */t/* in *Ute* oder */s/* in *lassen*).
- Palato-alveolar: Der vordere Teil der Zunge artikuliert im Übergangsbereich zwischen Alveolen und hartem Gaumen (z.B. */ʃ/* in *Maschen*).
- Palatal: Der Zungenrücken artikuliert mit dem harten Gaumen (z.B. */x/* in *Küche*).
- Velar: Der hintere Teil der Zunge artikuliert mit dem Gaumensegel (z.B. */k/* in *Kette*).
- Uvular: Der Zungenrücken artikuliert mit dem Zäpfchen (z.B. eine Ausspracheform von */R/* in *Rachen*).
- Glottal: Die Stimmbänder nähern sich einander an, um einen Verschluß (z.B. vor */a/* in *Axt*) oder eine Enge zu bilden (z.B. */h/* in *Hand*).

Bei der Artikulation von Konsonanten wird der Luftstrom in Rachen- und Mundhöhle von den Artikulationsorganen im wesentlichen auf vier Arten gehemmt (Artikulationsarten) ([Cry95]):

- Vollständiger Verschluß:
  - *Plosive*: An einem bestimmten Punkt des Vokaltrakts wird ein kompletter Verschluß gebildet, wobei das Gaumensegel angehoben ist. Der Luftdruck hinter dem Verschluß erhöht sich und wird dann schlagartig abgebaut (z.B. */p/* in *Paß*).
  - *Nasale*: An einem bestimmten Punkt im Mundraum wird ein kompletter Verschluß gebildet, wobei das Gaumensegel abgesenkt ist. Die Luft entweicht über die Nase (z.B. */m/* in *machen*).
  - *Affrikaten*: An einem bestimmten Punkt im Mundraum wird ein kompletter Verschluß gebildet; das Gaumensegel ist angehoben. Hinter dem Verschluß staut sich Luft an, doch wird dieser Überdruck relativ langsam abgebaut. Am Anfang hat der Laut deutlich plosiven Charakter, doch folgt darauf hörbare Reibung (z.B. */ts/* in *Katze*).
- Intermittierender Verschluß:
  - *Vibranten*: Ein Artikulationsorgan schlägt schnell und wiederholt gegen ein anderes, z.B. bei dem gerollten r in einigen Aussprachevarianten.
  - *Flaps*: Ein Artikulationsorgan schlägt nur ein einziges Mal gegen ein anderes (z.B. beim */d/* im englischen Wort *ladder*).
- Partieller Verschluß:
  - *Laterale*: An einer Stelle im Mund wird ein partieller Verschluß gebildet, und zwar so, daß der Luftstrom an den Rändern des Verschlusses entweichen kann (z.B. */l/* in *lassen*).
- Verengung:
  - *Frikative*: Zwei Artikulationsorgane nähern sich einander so weit, daß die durchströmende Luft hörbare Reibung erzeugt (z.B. */s/* in *lassen*, */ʃ/* in *Laschen* und */x/* in *lachen*).

1.3. Akustische Grundlagen

1.3.1. Der Sprachschall als Signal

Der Zweck des Sprechvorgangs ist die Übertragung von Information. Nach der Informationstheorie kann das Sprachsignal daher nach seinem Informationsgehalt beschrieben werden. Eine andere Möglichkeit ist die Beschreibung des informationstragenden Signals bzw. der akustischen Repräsentation durch den Sprachschall.

Die übertragene Information ist prinzipiell diskreter Natur. Jede Sprache bedient sich einer relativ geringen Anzahl von 30 bis 50 Phonemen (das sind die kleinsten bedeutungsunterscheidenden sprachlichen Einheiten). Vernachlässigt man gegenseitige Beeinflussungen benachbarter Phoneme und setzt man bei natürlicher Sprache eine mittlere Rate von 10 Phonemen pro Sekunde voraus, beträgt der tatsächlich benötigte Informationsfluß nur ca. 60 Bit/sec.
Für die digitale Repräsentation des Sprachschalls sind bei einer reinen Signalformcodierung (PCM) für eine ausreichende Silbenverständlichkeit üblicherweise 64000 Bit/sec erforderlich (ISDN-Standard), also mehr als das tausendfache. Erst mit modernsten technischen Verfahren der Quellencodierung, wie z.B. beim Digitalen Mobilfunk GSM, ist es gelungen, den erforderlichen Informationsfluß durch Entfernung von Irrelevanz und Redundanz auf ca. 2000 Bit/sec zu senken. Der immer noch vorhandene Überhang enthält u.a. sprecherspezifische Merkmale, wie den typischen Klang einer Stimme, und auch Nebeninformationen, wie z.B. den Gemütszustand des Sprechers (heiter oder depressiv) oder über die Geräuschumgebung (Fahrtwind, Raumechos, usw.).

Der folgende Abschnitt widmet sich ausschließlich den spezifischen akustischen Eigenschaften des Sprachsignals, die für Vokal- und Konsonantartikulation typisch sind. Hier soll im wesentlichen gezeigt werden, welche Unterschiede in der Zeit- und der Frequenzbereichsrepräsentation des Sprachsignals sichtbar sind.

1.3.2. Signalcharakteristika der Vokale

Vokale zeichnen sich durch ein periodisches Zeitsignal hoher Energie aus. Abbildung 1.6 zeigt das Oszillogramm für den Vokal */a/* mit der charakteristischen Wellenform jeder

Einzelschwingung.

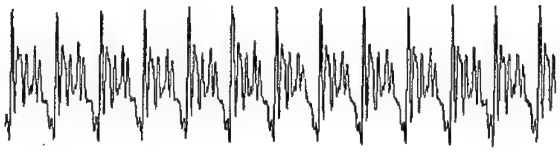

Abbildung 1.6.: Typisches Zeitsignal für den Vokal */a/*. Das Signal ist periodisch, die Wellenform jeder Einzelschwingung ist charakteristisch.

In der Frequenzbereichsdarstellung zeigen Vokalspektren typische Überhöhungen, die sogenannten Formanten. Die Frequenzlage der ersten beiden Formanten ist ein charakteristisches Merkmal, das u.a. bei der automatischen Spracherkennung herangezogen wird. Abbildung 1.7 zeigt ein Breitband-Spektrogramm für das Wort *Paß*.

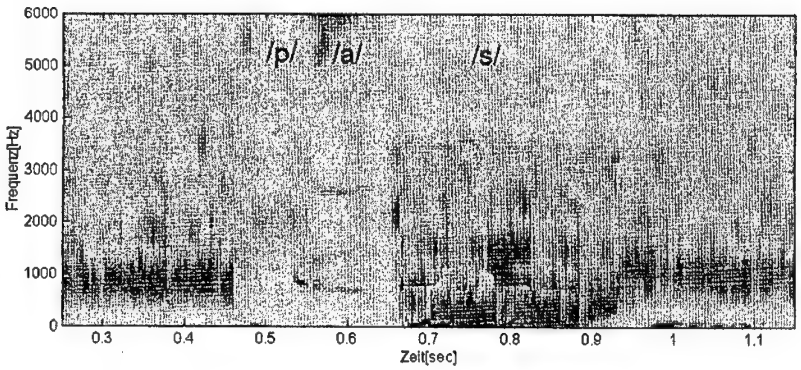

Abbildung 1.7.: Breitband-Spektrogramm für das Wort *Paß*. Der zentrale Vokal tritt deutlich durch seine typischen Formanten hervor.

Die Formanten treten bei dem zentralen Vokal deutlich als Bänder hervor. Für den Vokal */a/* nehmen die Formanten typisch Werte von ca. 700 Hz bzw. 1100 Hz an.

1.3.3. Signalcharakteristika der Konsonanten

Ein Teil der Konsonanten läßt sich akustisch in *plosive* und *frikative* Konsonanten differenzieren.
Plosive sind meist durch ein Intervall der Stille gekennzeichnet, bevor der Laut hervorgestoßen wird. Bei den stimmlosen Plosiven */p/,/t/,/k/* ist

- das Stilleintervall bis zur Verschlußlösung länger,
- die Intensität des Lautausstoßes stärker ausgeprägt und die
- Zeitspanne bis zum Einsatz der Stimmbandschwingung größer (*Voice-Onset-Time VOT*)

als bei den stimmhaften Plosiven */b/,/d/,/g/*. Abbildung 1.8 zeigt das Oszillogramm für den stimmlosen Plosiv */p/* in dem Wort *Paß*. Die Stimmbandschwingung setzt erst nach einiger Zeit ein.

Abbildung 1.8.: Typisches Zeitsignal für den Plosiv */p/* und den Übergang auf einen nachfolgenden Vokal. Die plötzliche Verschlußlösung bedingt einen starken Anstieg der Amplitude, und es verstreicht eine charakteristische Zeit (*VOT*), bis die Stimmbänder durch den pulmonalen Luftstrom zum Schwingen angeregt werden. Die Zeitskala ist identisch zu der in Abbildung 1.6.

Im Spektrogramm gem. Abbildung 1.7 ist eine weitere Besonderheit erkennbar. Im Zeitraum von der Verschlußlösung bis zum Einsatz der Phonation wird der Vokaltrakt durch den exspirierten pulmonalen Luftstrom auch ohne Phonation zu Resonanzen angeregt. Dadurch treten hier bereits Formantfrequenzen auf. Da sich die Artikulatoren

(in diesem Falle Lippen und Kiefer) bereits in der VOT auf die Zielstellung für den nachfolgenden Vokal zubewegen, gehen die Formanten stetig in die des Vokals über.

Frikative setzen ungeordnete Schallenergie oder Geräusche frei. Durch die stimmlose Anregung hat das korrespondierende Zeitsignal keine periodische Struktur, und es gleicht damit einem Rauschsignal. Abbildung 1.9 zeigt das resultierende Oszillogramm für den Frikativ */s/*.

Abbildung 1.9.: Typisches Zeitsignal für den Frikativ */s/*. Das Signal ist rauschartig, d.h., ohne jede periodische Struktur.

Das Spektrogramm gem. Abbildung 1.7 zeigt am auffälligsten die akustische Charakteristik von Frikativen, hier für den Frikativ */s/*. Bei vergleichsweise geringer Gesamtenergie liegt der Hauptenergieanteil bei sehr hohen Frequenzen oberhalb von ca. 3500 Hz.
Die spektralen Eigenschaften sind für jeden Frikativ unterschiedlich. Die Energie beim Frikativ */∫/* setzt beispielsweise bereits früher als bei */s/* ein, sie fällt dafür auf Grund der Lippenextrusion (Vorstülpung) zu hohen Frequenzen hin stärker ab.

1.4. Prosodische Merkmale

Vokale und Konsonanten sind die *Segmente* der gesprochenen Sprache, die zu Silben, Wörtern und Sätzen zusammengefügt werden. Während der Artikulation dieser Segmente schwankt die Aussprache jedoch in anderer Hinsicht; wir nutzen eine Vielzahl von stimmlichen und rhythmischen Veränderungen, die die Bedeutung des Gesagten auf verschiedenste Weise abändern (vgl. [Cry95]).
Die wichtigsten Effekte beruhen auf den psychoakustischen Eigenschaften von Lauten, das sind die *Tonhöhe* und die *Lautstärke*. Zusammen mit den Effekten, die sich aus

Sprechtempo und -rhythmus ergeben, werden sie als *prosodische* Merkmale bezeichnet.

Sämtliche Sprachen scheinen die Unterscheidung zwischen fallender und steigender Tonhöhe zu nutzen, um zwischen Aussagen und Fragen zu differenzieren. Zusätzlich lassen sich noch viele weitere Töne nutzen, um verschiedene Nuancen und Betonungen auszudrücken ([Cry95]).

Die Lautstärke verdeutlicht grobe Bedeutungsunterschiede, aber auch feine Kontraste zwischen den verschiedenen Silben eines Wortes. Die Lautstärke von Silben wird meist als *Betonung* bezeichnet, wobei eine Silbe *betont* oder *unbetont* sein kann. Der Begriff *Akzent* wird dann verwendet, wenn das Hervorheben einer Silbe von Tonhöhe und Lautstärke abhängt ([Cry95]).

Die Schwankungen im Sprechtempo sind eine dritte suprasegmentale Größe. Die Anzahl der produzierten Silben, Wörter und Sätze innerhalb einer bestimmten Zeit kann verschiedene Bedeutungen vermitteln. Ein besonders schnell gesprochener Satz drückt beispielsweise Dringlichkeit aus, während ein reduziertes Sprechtempo Überlegung oder Betonung vermittelt. Einzelne, abgehackte Silben können hingegen Verärgerung signalisieren.

Tonhöhe, Lautstärke und Sprechtempo ergeben zusammen den *Rhythmus* einer sprachlichen Äußerung.

2. Neurogene Sprechstörungen und deren klinische Diagnostik

Die Störungen zentralmotorischer Programmierungs- und Steuerungsprozesse des Sprechbewegungsablaufs werden unter dem Begriff der *neurogenen Sprechstörungen* zusammengefaßt ([Zie85]).
Ausgenommen sind dabei also alle Störungen, die durch lokale Veränderungen an den beteiligten Organen verursacht werden (z.B. Stimmbandknötchen, Kiefer-Gaumen-Spalten, Makroglossie etc.), die Sprachentwicklungsstörungen (einschließlich des angeborenen Stottersyndroms und der Sprechstörungen bei Gehörlosen) und die psychogenen (d.h. die Intelligenz oder Aufmerksamkeit betreffenden) Sprechstörungen ([Zie85]).
Ausgenommen sind auch die Sprechstörungen, die auf einer Schädigung peripherer, die Sprechmuskulatur innervierender Neurone bei intakter zentralnervöser Steuerung beruhen (z.B. Rekurrensparese, periphere Facialisparese etc.) ([Zie85]).
Eine weitere Abgrenzung muß zu den Beeinträchtigungen „höherer"Prozesse der Sprachverarbeitung, den *Aphasien*, getroffen werden: bei den aphasischen Syndromen erstreckt sich die Störung (wenn auch in unterschiedlicher Ausprägung) auf alle expressiven (d.h. Sprechen und Schreiben) und rezeptiven (d.h. Verstehen und Lesen) Modalitäten, während bei den neurogenen Sprechstörungen ausschließlich die mündliche Sprachproduktion in ihren motorischen Aspekten beeinträchtigt ist. Eine klare Differenzierung zwischen sprechmotorischen und sprachsystematischen Störungen der Lautproduktion bei aphasischen Patienten ist jedoch nicht immer leicht möglich ([Zie85]).

Unter den neurogenen Sprechstörungen differenziert man die

- *Dysarthrien*, die alle Störungen der Kontrolle von Kraft, Bewegungstempo und

Bewegungsumfang bei der *Ausführung von Sprechbewegungen* umfassen, und die

- *Sprechapraxie*, die als Störung der *Programmierung von Sprechbewegungen* gilt ([Zie93b]).

2.1. Ätiologien

Beeinträchtigungen der zentralnervösen Steuerung des Sprechbewegungsapparates können, wie die übrigen zentral-motorischen Störungen, bedingt sein durch

- entzündliche oder degenerative Prozesse,
- Neoplasmen (Geschwulste oder Tumore),
- vaskuläre Erkrankungen (Gefäßerkrankungen),
- Schädel-Hirn-Traumen (Verletzungen durch Gewalteinwirkung).

Die häufigste Ursache bildet die Gruppe von Erkrankungen, die unter dem Sammelbegriff *Schlaganfall* zusammengefaßt wird. Hierzu gehören Hirninfarkte (vaskuläre zerebellare Insulte) und innere Blutungen des Gehirns.
Die zweithäufigste Ursache mit ca. 20000 Fällen pro Jahr ([Mum90]) wird durch die Gruppe der *Schädel-Hirn-Traumen* gebildet, vor allem bedingt durch Verkehrsunfälle. Die Angaben zur Auftretenshäufigkeit dysarthrischer Störungen bei Patienten mit schwerem, gedecktem Schädel-Hirn-Trauma schwanken zwischen 30 und 50 Prozent ([Gilc79],[Sar86],[Zie87],[Pro88]).
Die bedeutendste degenerative Erkrankung des zentralen Nervensystems ist das *Parkinson-Syndrom*, bei dem in 80 Prozent aller Fälle mit einer Beeinträchtigung von Sprechatmung, Stimme und/oder Artikulation gerechnet werden muß (vgl. [Ack89]).
Von den an *Multipler Sklerose* erkrankten Patienten sind nach amerikanischen Schätzungen mehr als die Hälfte von Sprechstörungen betroffen ([Dar85]).
Bei den selteneren Erkrankungen wie der *Chorea Huntington*, der *myatrophen Lateralsklerose (ALS)* oder den degenerativen Erkrankungen des Kleinhirns kommt ebenfalls in der Mehrheit aller Fälle eine Beeinträchtigung der Sprechmotorik vor ([Hus73],[Gilm81],[Dar85]).

2.2. Syndrome und Störungsprofile

2.2.1. Dysarthrien

Dysarthrien sind erworbene neurogene Sprechstörungen. Sie werden durch eine Schädigung des zentralen oder des peripheren Nervensystems verursacht und beruhen auf einer Beeinträchtigung der Steuerung und Ausführung von Sprechbewegungen [Zie98b].

Dysarthrische Störungen entstehen durch Schädigung neuronaler Strukturen, die an der Steuerung der Bewegungen von

- Atmungsmuskulatur,
- Kehlkopf,
- velopharyngealer Muskulatur,
- Zunge,
- Kiefermuskulatur und
- Lippen

beteiligt sind. Es handelt sich daher um eine *sprechmotorische* Störung, genauer um eine Beeinträchtigung „elementarer"Prozesse der Bewegungsausführung, nicht etwa um Störungen von Prozessen der Planung oder Programmierung von Sprechbewegungen.
Es lassen sich mehrere Dysarthrieformen unterscheiden, denen jeweils Schädigungsmechanismen mit zugehörigen prototypischen Beschreibungen der Syndrome nach auditiven Merkmalen zugrunde liegen. Die nachfolgende Differenzierung ergibt sich nach der Art der Bewegungsstörung.

Schlaffe Parese: Bei der schlaffen Parese tritt eine Minderung der Kraft und ein Erschlaffen der betroffenen Muskulatur ein. Dies kann auch zu isolierten Störungen der Atmung, Stimme oder Artikulation führen. Beobachtet wird u.a. eine Beeinträchtigung der Adduktionsfähigkeit der Stimmlippen, eine verlangsamte und wenig differenzierte Zungenbewegung mit verringerter Bewegungsamplitude sowie eine Kieferheberschwäche

mit verlangsamter Artikulation und unvollständigem Mundschluß.
Auditiv werden bei der schlaffen Dysarthrie Merkmale wie eine verkürzte Exspiration, eine behauchte und rauhe Stimmqualität und eine reduzierte Lautstärke bei erniedrigter Stimmlage beschrieben. Die Artikulation ist unscharf und hypernasal. Das Sprechen ist verlangsamt bei monotoner Intonation und vermehrten Sprechpausen (vgl. [Zie98b]).

Spastische Parese: In der klinischen Diagnostik der Gliedmaßenmotorik wird Spastizität im allgemeinen als ein kombiniertes Störungsmuster bestehend aus erhöhtem Muskeltonus, gesteigerten Eigenreflexen und gestörter Feinmotorik verstanden. Dieses Konzept der Spastizität ist nur bedingt auf die Sprechmotorik übertragbar. Dennoch gibt es eine Reihe von Anzeichen, die für das Vorliegen eines der Spastizität vergleichbaren Syndroms bei dysarthrischen Patienten sprechen:
Die willkürliche Beweglichkeit der einzelnen am Sprechbewegungsablauf beteiligten Organe ist eingeschränkt oder aufgehoben, eine Tonussteigerung der Muskulatur läßt sich aus der sichtbaren Kontraktion erschließen („zigarrenförmige"Zunge, verkürzte oder vedickte Stimmbänder), und die reflektorische Beweglichkeit ist erhalten.
Für das auditive Störungsbild sind Artikulationsunschärfe, Verlangsamung, Hypernasalität und eine monotone Sprechweise charakteristisch. Während bei der schlaffen Parese eine behauchte Stimmqualität bei erniedrigter Stimmlage wahrgenommen wird, ist bei der spastischen Parese die Stimmqualität gepreßt bei insgesamt eher erhöhter Stimmlage (vgl. [Zie98b]).

Rigidität: Die betroffenen Gliedmaßen bauen hier einen beschleunigungsunabhängigen Widerstand auf und verharren danach in der eingenommenen Position. Die Rigidität ist eines der Kardinalsymptome der Parkinson'schen Erkrankung. Da sich der Beitrag einer rigiden Komponente zum Störungsbild der Parkinson-Dysarthrie nur schwer von anderen Komponenten der Parkinson'schen-Bewegungsstörung trennen läßt, sind die Merkmale dieser Sprechstörung mit den hypokinetischen Merkmalen der Parkinson-Dysarthrie zu einem *rigid-hypokinetischen* Syndrom zusammengefaßt (vgl. [Zie98b]); eine Erläuterung hierzu folgt.

Akinesie: Der Begriff Akinesie charakterisiert die klinische Beobachtung eines Fehlens oder einer deutlichen Reduktion von Willkürbewegungen, wobei die betroffene Muskulatur nicht paretisch (d.h. nicht gelähmt) ist. Dieses Symptom ist typisch für Parkinsonismus ([Zie98b]).
MARSDEN faßt unter dem Begriff der Akinesie drei Komponenten zusammen: Eine Störung der Bewegungsinitiierung, eine Reduktion des Bewegungsumfangs (*Hypokinesie*) und eine Verlangsamung von Bewegungen (*Bradykinesie*) ([Mars89]).
Das Dysarthriesyndrom, das für Parkinsonkranke charakteristisch ist, wurde von DARLEY als *hypokinetische Dysarthrie* bezeichnet ([Dar75]). Nachdem die hypokinetische und die (bei Parkinsonismus ja ebenfalls vorhandene) rigide Komponente der Sprechbewegungsstörung auditiv nicht auseinanderzuhalten sind, wurde der Begriff der *rigid-hypokinetischen Dysarthrie* eingeführt ([Zie98b]).
Die betroffenen Patienten sprechen typischerweise leise, die Sprechstimmlage ist aufgrund des erhöhten Tonus der Stimmlippen häufig zu hoch. Die Stimme ist behaucht und rauh. Sie ist wenig modulationsfähig und klingt monoton. In ausgeprägteren Fällen ist die Artikulationsschärfe reduziert (vgl. [Zie98b]).

Dyskinesien: Unter diesem Begriff werden unterschiedliche Formen unwillkürlicher Muskelaktivation zusammengefaßt. Bei der *choreatischen* Form treten Sprechstörungen auf, die auditiv durch unwillkürliche Unterbrechungen des Redeflusses, abrupte Veränderungen von Stimmqualität, Tonhöhe oder Lautstärke und verlangsamte, zeitweise „explosive"Artikulation ([Ram86]) charakterisiert sind.
Bei den *dystonischen* Formen werden Sprechstörungen wie intermittierende Artikulationsstörungen, gepreßte und rauhe Stimmqualität sowie Stimmabbrüche beobachtet.
Bei der Form der *Myoklonie* treten beim Sprechen periodische Schwankungen von Tonhöhe und Lautstärke auf.

Ataxie: Patienten mit Ataxien im Rahmen eines zerebellären Syndroms zeigen Störungen der Bewegungskoordination, Intentionstremor, Bewegungsverlangsamung sowie Gang- und Standunsicherheit.
Die auditiven Leitsymptome der ataktischen Dysarthrie ergeben sich aus dem Charakter der ataktischen Bewegungsstörung. Es kann zu inadäquaten Einatmungspausen, in ausgeprägtem Fall auch zu inspiratorischem Sprechen kommen; die Stimmqualität und

die Tonhöhe und Lautstärke variieren. Auch die artikulatorischen Störungen fluktuieren; Phasen mit reduzierter Artikulationsschärfe können mit „explosiver", stark fortisierter Artikulation wechseln. Die Artikulation ist insgesamt verlangsamt und durch Lautdehnungen oder verlängerte Pausen charakterisiert. Es kommt zu Verlangsamungen und Auffälligkeiten in der Zeitstruktur auf Satzebene (vgl. [Zie98b]).

Tremor: Unter diesem Begriff werden unwillkürliche rhythmisch-oszillatorische Bewegungen der Extremitäten, des Kopfes oder Rumpfes oder der am Sprechen beteiligten Muskulatur zusammengefaßt. Der *Stimmtremor* ist durch rhythmische Oszillationen von Tonhöhe und Lautstärke charakterisiert. Er kann aus einem Tremor unterschiedlicher Muskelgruppen resultieren (vgl. [Zie98b]).

Störungen des Redeflusses: Erworbene neurogene Störungen des Redeflusses (z.B. „erworbenes Stottern") werden üblicherweise nicht als dysarthrische Störungen gesehen. Allerdings wird nach heutiger Auffassung das während der Sprachentwicklung auftretende Stottersyndrom eindeutig als sprechmotorische Störung interpretiert ([Hul98]). Daher sollte auch die erworbene neurogene Form dieses Syndroms im Zusammenhang mit den Dysarthrien diskutiert werden (vgl. [Zie98b]).

2.2.2. Sprechapraxie

Die Sprechapraxie wird in der Literatur als eine *Störung der Programmierung von Sprechbewegungen* gekennzeichnet.
Gegenüber den aphasisch bedingten phonologischen Störungen, welche gemeinhin als Defizite abstrakter sprachlicher Prozesse und Konzepte, insbesondere der Selektion und Sequenzierung von Phonemen, betrachtet werden, findet eine Abgrenzung statt, da diese vermutlich auf einer höheren Verarbeitungsebene anzusiedeln sind ([Zie91]).
Das derzeitige Sprechapraxie-Konzept ist jedoch nicht unumstritten, weil die Isolierung von sprechmotorischen Programmierungsprozessen gegenüber anderen Komponenten innerhalb des Gesamtprozesses der Transformation von Wörtern und Sätzen in Sprechbewegungen nur bedingt möglich ist. Ungeachtet dessen steht die klinische Bedeutung des als Sprechapraxie bezeichneten Störungsbildes außer Frage.
Sprechapraktische Störungen finden sich auf der Ebene der Lautbildung (segmentale

Ebene), der Prosodie (suprasegmentale Ebene) und der Ebene der linguistisch nicht beschreibbaren besonderen Merkmale beim Sprechen (Sprechverhalten).

Auf segmentaler Ebene werden hierbei folgende Symptome differenziert:

- Phonetische Entstellungen: Der betreffende Laut wird nicht korrekt gebildet, jedoch kann er immer noch der Kategorie des Ziellautes zugeordnet werden. Es werden Längungen von Lauten, Denasalierung nasaler Konsonanten, Entstimmung von stimmhaften Konsonanten, übermäßige Behauchung stimmloser Plosive, unscharfe Frikativrealisierung oder eine Vor- und Rückverlagerung lingualer Konsonanten (vgl. [Ode90],[Zie86],[Vog88]) beobachtet.
- Phonematische Paraphasien: Der Patient läßt einen Laut aus, fügt einen zusätzlichen Laut hinzu oder er realisiert anstelle des vorgesehenen Ziellauts einen alternativen Laut, den er hingegen korrekt artikuliert.
- Mischformen phonetischer Entstellungen und phonematischer Paraphasien: Anstelle des erwarteten Phonems wird eine andere Lautkategorie realisiert, diese aber durch Fehlartikulation phonetisch entstellt.

Phonematische Fehler sind nicht nur bei Sprechapraxiepatienten, sondern auch bei Aphasikern ohne Sprechapraxie beobachtbar. Lautentstellungen sind dagegen für dysarthrische Patienten typisch. Im Falle der Dysarthrie sind diese Fehler jedoch konstant und somit in gewisser Weise vorhersagbar. Dies gilt jedoch nicht für die Sprechapraxie, da hier die auftretenden Fehler variabel und unvorhersagbar sind ([Zie89]).
Man spricht in diesem Zusammenhang von der für Sprechapraxie typischen *Inkonstanz* und *Inkonsistenz* von Fehlern.

Auf suprasegmentaler Ebene werden Störungen der Prosodie beschrieben. Ein markantes Merkmal hierfür ist das silbische (abgehackte) Sprechen, dessen Eindruck vor allem durch intersilbische Pausen, die Betonung oder Längung von Nebensilben oder eine mangelhafte Koartikulation an Silbengrenzen entsteht. Der Intonationsverlauf ist durch Sprechpausen oder Iterationen von Lauten oder Silben beeinflußt, die zeitliche Struktur von Wörtern und Sätzen hingegen durch Lautdehnungen und -auslassungen gestört.

Das Sprechverhalten ist bei den meisten sprechapraktischen Patienten von suchenden Bewegungen der Artikulationsorgane geprägt (*Suchverhalten*). Meist ist dieses Suchen mit Sprechpausen verbunden. Es tritt am häufigsten bei der Initiierung eines Wortes oder Satzes auf. Zum Teil verlaufen diese Suchbewegungen stumm, zum Teil sind sie aber auch hörbar. Die so produzierten Laute können sogar teilweise einer Phonemkategorie zugeordnet und so als ergänzende Laute interpretiert werden, wodurch sich ein Brückenschlag zu den segmentalen Fehlern ergibt.
Daneben läßt sich das Merkmal der *Sprechanstrengung* beobachten. Es ist aber nur schwer faßbar, da auch andere Verhaltensweisen des Patienten wie ein starkes Suchverhalten und der hieraus resultierende unflüssige Sprechfluß oder eine erhöhte Sprechstimmlage auf eine Sprechanstrengung hindeuten könnten (vgl. [Tei95]).

2.3. Untersuchungsverfahren

Der Hauptteil der Diagnostik, soweit sie in der Verantwortung von Sprachtherapeuten liegt, beruht auf einer auditiven Analyse der Äußerungen des Patienten. Streng genommen handelt es sich hierbei nicht ausschließlich um einen Hörbefund, da einige wichtige Zusatzinformationen auch gesehen oder getastet werden können. Der Nachteil einer Diagnostik auf perzeptiver Grundlage liegt jedoch auf der Hand:

- Die Befunde sind subjektiv und ihre Qualität hängt von der Schulung und Erfahrung des Untersuchers ab.
- Aus den hörbaren Merkmalen kann nur in sehr unzulänglicher Weise auf die zugrundeliegende Bewegungsstörung geschlossen werden.

Diese Schwächen können unter anderem durch Verwendung gezielter diagnostischer Aufgaben und eine Vereinfachung der Beurteilungskriterien einigermaßen wettgemacht werden (vgl. [Zie98b]).

2.3.1. Verfahren in der Dysarthriediagnostik

2.3.1.1. Auditive Verfahren

Die auditive Diagnostik läßt sich in die Allgemeine Beobachtung, die Analyse phonetischer Störungsmerkmale und die Analyse der Verständlichkeit differenzieren. Die Beurteilung der Sprechfähigkeit erfolgt zum einen anhand der spontanen Äußerungen eines Patienten im Rahmen eines Gesprächs, z.B. des Anamnesegesprächs, zum anderen anhand von Aufgabenstellungen, bei denen der Patient spezifische Äußerungen produzieren (meist nachsprechen) soll.

Allgemeine Beobachtung: Die Allgemeine Beobachtung bezieht sich hier auf die Körperhaltung des Patienten und auf sichtbare Störungsmerkmale wie pathologischer Atmungstyp, Störungen von Tonus und Bewegungsfähigkeit der Lippen und des Kiefers, und auf Dyskinesien, auf ataktische Zeichen oder einen Tremor der Gesichts-, Rumpf- und Extremitätenmuskulatur (vgl. [Zie98b]).

Analyse phonetischer Störungsmerkmale: Bei der Analyse phonetischer Störungsmerkmale steht die funktionsorientierte Diagnostik der Sprechstörung nach den Funktionskreisen der Sprechatmung, der Phonation und der Artikulation im Vordergrund. Jede dieser Störungskomponenten kann dabei anhand einer Auswahl von charakteristischen Symptomen beschrieben werden.
Das Vorliegen dieser Symptome wird einerseits anhand der spontanen Äußerung des Patienten ermittelt, zum anderen gezielt mit Hilfe von spezifischem Sprachmaterial geprüft. Schließlich kann es auch sinnvoll sein, in „Maximalleistungsaufgaben"die Leistungsgrenzen des Patienten auszuloten ([Zie98b]). Die Analyse der Spontansprache liefert zweifellos die alltagsrelevantesten Aussagen, während die Beurteilung von speziell ausgewähltem Sprachmaterial meist eine höhere diagnostische Spezifität ermöglicht. Maximalleistungsaufgaben sind in der Regel sensitiver als sprachliche Aufgaben, wegen ihres nichtsprachlichen Charakters lassen sie jedoch nur bedingt Rückschlüsse auf die Sprechstörung zu ([Zie98c]).

Der Grad und das Ausmaß einer *Sprechatmungsstörung* kann anhand verschiedener Merkmale zur Beschreibung eines pathologischen Atmungstyps, respiratorischer Insuffizienz und gestörter Lautstärkekontrolle manifestiert werden. Da die Einatmungshäufigkeit in der Spontansprache nur schwer zu beurteilen ist, werden in der Praxis häufig auch Zählaufgaben verwendet, d.h., der Patient zählt in „zügigem"Tempo von 1-20. Gesunde Personen benötigen bei dieser Aufgabe selten mehr als eine Zwischenatmung. Die Einatmungspausen lassen sich in ein vorgefertigtes Schema eintragen und auszählen. Eine weitere Aufgabe ist die Lauthalteaufgabe, bei der der Patient gebeten wird, einen vorgegebenen Laut so lange wie möglich anzuhalten. Die Haltedauer wird mit einer Stoppuhr oder im Spracheditor eines Computers gemessen und mit Normwerten verglichen. ZIEGLER und seine Mitarbeiter ermittelten in einer Untersuchung an 200 Dysarthriepatienten bei dieser Vorgehensweise in 58 Prozent der Fälle verkürzte Vokalhaltedauern und in 68 Prozent der Fälle verkürzte Frikativhaltedauern. Der Einfluß eines velopharyngealen Luftverlusts läßt sich bei Lauthalteaufgaben durch eine Nasenklammer feststellen (vgl. [Zie98b]).

Zur Untersuchung einer *Stimmstörung* stehen Merkmale zur Beurteilung der Stimmqualität (rauh, behaucht, gepreßt), der Stimmstabilität (Tonhöhen- und Lautstärkeschwankungen, Stimmschwund, Stimmzittern oder Stimmabbruch) und von Verschiebungen der Stimmlage (zu hoch, zu tief, zu laut oder zu leise) zur Verfügung. Durch Nachsprechaufgaben können hierbei spezifische Anforderungen an die Einstellbewegung der Stimmlippen beim Stimmeinsatz geprüft werden.
Eingesetzt werden auch häufig Maximalleistungsaufgaben wie die Aufgabe der maximalen Vokalhaltedauer, die beispielsweise Aufschluß über die Vollständigkeit der Glottisadduktion geben kann. Einige Stimmqualitätsmerkmale wie Tonhöhenschwankungen, Lautstärkeschwankungen oder Stimmzittern sind zudem anhand gehaltener Vokale leichter zu diagnostizieren als anhand der Spontansprache. Zur Prüfung des Tonhöhenumfangs werden Tonhöhenvariationsaufgaben eingesetzt (vgl. [Zie98b]).

Bei der Untersuchung der *Artikulationsstörung* werden die einzelnen beteiligten Subsysteme wie z.B. Gaumensegel, Lippen und Zunge möglichst selektiv überprüft. Dabei sind vor allem die Bewegungen von Lippen und Kiefer, zum Teil auch der Zunge, visuell beurteilbar. Auditiv werden anhand der Spontansprache Merkmale zur

Beschreibung der Kieferöffnung, der Konsonantartikulation, der Vokalartikulation, abweichender Artikulationsstellen oder der Nasalität verwendet.
In Nachsprechaufgaben kann man sich eine phonematische Systematik des eingesetzten Vokal- und Konsonantinventars zunutze machen, um spezifische artikulatorische Funktionen gezielt zu untersuchen. Es liegt nahe, hier Sätze und Wörter mit einer Häufung von Sprachlauten vorzugeben, die vorwiegend durch ein bestimmtes artikulatorisches Organ gebildet werden. Um artikulatorische Abläufe an der Leistungsgrenze zu untersuchen, sind Maximalleistungsaufgaben in der Form *Schneller Silbenwiederholungen* bzw. artikulatorischer *Diadochokinese* verbreitet. Dabei muß eine vorgegebene Silbe- oder Silbenfolge so rasch wie möglich wiederholt werden. Im einfachsten Fall wird dabei die Zahl der Silben innerhalb eines festen Zeitintervalls bestimmt, wodurch jedoch nur die erreichte Silbenrate, nicht jedoch Unregelmäßigkeiten berücksichtigt werden (vgl. [Zie98b]).

Störungen der *Prosodie* können anhand von Merkmalen zur Beschreibung von Sprechtempo (verlangsamt oder beschleunigt), Redefluß (Lautdehnungen, Pausen, Iterationen), Rhythmus/Akzent (Betonung von Nebensilben, silbisches Sprechen) und Intonation (monotones Sprechen) beschrieben werden. Da Sprechtempo und Redefluß bei freiem Sprechen auch durch Sprachproduktionsprobleme auf der Ebene vorsprachlicher kognitiver Prozesse oder auf verschiedenen Ebenen der sprachlichen Enkodierung gestört sein können, sollte dieser Aspekt der Prosodie zusätzlich durch Nachsprech- oder Leseaufgaben geprüft werden ([Zie98b]).

Durch die beschriebenen qualitativen Merkmale und Beurteilungskriterien ergibt sich - zusammen mit den Ergebnissen verschiedener Zusatzuntersuchungen - ein umfassendes Bild von der Gesamtstörung. Sieht man von den beschriebenen Maximalleistungsaufgaben ab, so enthält diese Beschreibung allerdings noch keine quantitativen Aussagen über das Ausmaß der Störung in den verschiedenen Bereichen. Dazu werden die Beurteilungsmerkmale auf geeigneten Skalen bewertet (vgl. [Zie98b]). Das bekannteste Beispiel solcher Skalen ist das von DARLEY und Mitarbeitern ([Dar75]) entwickelte auditive Beurteilungssystem, welches jedoch auf Grund seiner nicht hinreichenden Objektivität nicht unumstritten ist. Alternativ wurde daher von ZIEGLER ([Zie98b]) die Verwendung eines Systems mit einer verringerten Anzahl und weniger differenzierten Skalen als nach DARLEY vorgeschlagen, deren einzelne Ausprägungsniveaus operational beschrieben werden.

Analyse der Verständlichkeit: Die bisher beschriebene Diagnostik erlaubt noch keine Rückschlüsse auf die hieraus resultierenden kommunikativen Einschränkungen des Patienten. Die *Verständlichkeit* zählt wohl zu den wichtigsten Kriterien.
Für eine Verständlichkeitsmessung werden drei Gruppen von Verfahren unterschieden: Schätzskalen, Transkriptionsverfahren und Wortidentifikationsverfahren.

Die Verständlichkeitsskala des *National Institute for the Deaf* entstammt dem Bereich der Diagnostik von Sprechstörungen hörgeschädigter Patienten. Die Ausprägungsstufen werden auf einer fünfstufigen Skala von „unverständlich"bis „völlig verständlich"beschrieben. Trotz psychometrischer Unzulänglichkeiten besitzt diese Skala eine vergleichsweise hohe Reliabilität (vgl. [Zie98b]).

Bei den Transkriptionsverfahren werden die vom Patienten gesprochenen Äußerungen durch den Untersucher Wort für Wort transkribiert. Die Zahl der falsch oder nicht transkribierten Wörter dient dabei als Verständlichkeitsmaß ([Yor84]). Da die in einem sinnvollen Satz enthaltenen Wörter aufgrund des Kontextes unterschiedlich gut vorhersagbar sind, hängt das Resultat entscheidend von der Struktur des gewählten Sprachmaterials ab. Für die klinische Standarddiagnostik erscheint diese Klasse von Verfahren daher weniger gut geeignet ([Zie94]).

Bei den Wortidentifikationsverfahren muß der Beurteiler die Wörter einer vom Patienten gesprochenen Wortliste unter einer Auswahl von ähnlich klingenden Alternativen identifizieren. Der Verständlichkeitswert ergibt sich aus der Anzahl der korrekt identifizierten Wörter. Das Münchner Verständlichkeitsprofil ([Zie93a]) hat sich hier als reliabel und valide für die Verständlichkeitsbeurteilung dysarthrischer Patienten herausgestellt. Es erfordert für die Durchführung ein speziell dafür entwickeltes Computerprogramm ([Ahr92]).

2.3.1.2. Akustische Verfahren

Eine Analyse der Arbeitsmethoden in der klinischen Diagnostik neurogener Sprechstörungen zeigt, daß hier - wie bereits beschrieben - vorwiegend auditive Verfahren zum Einsatz kommen, d.h., ein geschulter Hörer beurteilt einen Patienten nach seinem persönlichen Höreindruck. Solche Hörersitzungen sind zeitintensiv und subjektiv.

Akustisch-phonetische Verfahren, also der unterstützende Einsatz technischer Systeme, tragen hier zu einer Objektivierung des klinischen Bildes bei. Im Gegensatz zum Höreindruck besteht hier insbesondere die Möglichkeit, Verläufe von Sprechstörungen zu beschreiben, da subjektive Wahrnehmungseffekte ausgeschlossen sind.
Die akustische Analyse liefert eine physikalische Beschreibung des Sprachschalls auf einer vermittelnden Ebene zwischen den Bewegungsabläufen und der wahrgenommenen Sprache. Bei der Analyse des Sprachsignals lassen sich zeitliche, spektrale und durch mathematische Berechnungen gewonnene Parameter erfassen. Von Interesse sind hier primär die zeitlichen Verläufe von Schalldruck, Intensität und Grundfrequenz sowie spektrale Energieverteilungen.

Aus der Grundfrequenz lassen sich spezielle Variationsmaße für die Stimmstabilität extrahieren: Jitter und Shimmer. Jitter bezieht sich auf die Variabilität der Dauer aufeinanderfolgender Grundperioden, Shimmer hingegen auf die Variabilität der Amplitude aufeinanderfolgender Grundperioden. Der Bezug zu bestimmten gestörten physiologischen Abläufen an den Stimmbändern kann nicht eindeutig hergestellt werden. Ebenso gelingt die eindeutige Zuordnung zu bestimmten perzeptiven Stimmerkmalen nur zu einem bestimmten Grad ([Zie98b]).

Die spektrale Energieverteilung läßt sich zur Bestimmung charakteristischer Frequenzbänder mit besonders großer Signalintensität oder Resonanz heranziehen (Formanten). Aus der Darstellung der beiden ersten Formanten lassen sich charakteristische Abstände der Frequenzbänder für die Realisierung unterschiedlicher Vokalqualitäten bestimmen (siehe [Alf82]); somit sind Rückschlüsse auf die Vokalartikulation möglich. Ebenso lassen sich charakteristische spektrale Merkmale für Frikative und Nasale finden.

Die Nasometrie ist ein akustisches Verfahren zur Nasalitätsbestimmung, bei dem sich durch Verwendung zweier getrennter Mikrophone die nasalen und oralen Anteile des Sprachsignals differenziert beurteilen lassen. Dabei sind Nasen- und Mundraum durch eine schallisolierende Platte getrennt. Der Verlauf des relativen Anteils nasaler Energie (*Nasalanz*) ist ein Maß für die Nasalität.

Obwohl im Prinzip bereits die notwendigen Methoden und Verfahren der akustischen Analyse verfügbar sind und eine breite klinische Nutzung befürwortet und auch

angestrebt wird, finden sich bis heute nicht einmal die einfachsten akustischen Parameter routinemäßig in klinischen Befunden und Berichten ([Merk99b]). Das in Kooperation der Universität der Bundeswehr München und der Entwicklungsgruppe klinische Neuropsychologie entwickelte und im Rahmen dieser Arbeit vorgestellte MODIAS-System ist ein Versuch, diese Situation zu verändern. Mit MODIAS wurde erstmals ein klinisch nutzbares Diagnostik-System für die Standarduntersuchung neurogener Sprechstörungen und die Befunderstellung anhand akustischer Parameter zur Verfügung gestellt. Neben der Berechnung leicht interpretierbarer und diagnostisch relevanter akustischer Parameter bietet es den Vorteil, den Untersucher durch ein umfassendes und vollständiges Untersuchungsprotokoll zu führen. Das System ist speziell für den klinischen Einsatz mit seinen sehr spezifischen Anforderungen konzipiert. MODIAS ist bereits bei verschiedenen klinischen Einrichtungen erfolgreich im Einsatz.

2.3.2. Verfahren in der Sprechapraxiediagnostik

Zur Beurteilung der Sprechfähigkeit eignen sich spontane Äußerungen des Patienten und Nachsprechaufgaben. Bei einer teilweisen oder völligen Unfähigkeit zur Sprachproduktion (Mutismus) muß zunächst die Ursache hierfür geklärt werden.

Ein Nachsprechtest sollte dabei folgende Anforderungen erfüllen:

- Alle Phoneme des Deutschen und die wichtigsten Konsonantverbindungen im An- und Auslaut sollten abgedeckt sein.
- Testwörter sollten ein- bis zweisilbige, konkrete, hochfrequente, ungebeugte Nomina sein, um lexikalische Effekte sowie Wortlängeneffekte gering zu halten.
- Um eine mögliche Fehlerinkonstanz und -inkonsistenz zu prüfen, sollte ein gewisser Anteil der Testwörter mehr als einmal, statistisch gestreut, präsentiert werden.

Eine eventuell beobachtete Sprechanstrengung sowie das Auftreten von Suchbewegungen muß protokolliert werden. Fehler der segmentalen und suprasegmentalen Ebene werden durch Transkription der Äußerung erfaßt (vgl. [Tei95]).

Folgende Kriterien führen differentialdiagnostisch zum Ausschluß einer Sprechapraxie (nach [Zie98b]):

- Das Auftreten konstanter phonetischer Störungsmerkmale (z.B. Lenisierung, Hypernasalität etc.).
- Das Bestehen einer der artikulatorischen Störung annähernd vergleichbaren Dysphonie.
- Das Fehlen phonematischer Zeichen (z.B. Phonemsubstitutionen).
- Das Fehlen von Suchbewegungen.

Während die Analyse segmentaler Defizite nach auditiven Kriterien ein klinisch etabliertes Verfahren darstellt, gibt es für die Quantifizierung der Zeitaspekte gestörter Satzproduktion noch keine klinisch anwendbaren Verfahren ([Merk97a]).

Das hier vorgestellte MODIAS-System stellt erstmals diese Funktionalität zur Verfügung. Es erlaubt die akustische Analyse des Zeitmusters der Satzproduktion bei gleichzeitiger auditiver Analyse nach phonetischen und phonematischen Fehlern. Es eignet sich daher u.a. zur Differenzierung unterschiedlicher Störungsmuster und zur Analyse der Faktoren, die das Auftreten von Unflüssigkeiten beeinflussen.

3. MODIAS: Diagnostisches und Technisches Systemkonzept

Dieses Kapitel stellt das *Modulare Diagnostik-System für Sprechstörungen MODIAS* vor, das im Zeitraum von 1994 bis 1999 in einer Kooperation des Instituts für Nachrichtentechnik der Universität der Bundeswehr München und der Entwicklungsgruppe Klinische Neuropsychologie EKN des städtischen Krankenhauses München-Bogenhausen entwickelt wurde. Nach einer Analyse des klinisch/therapeutischen Bedarfs werden hier zunächst die generellen Anforderungen an ein praktikables technisches Hilfsmittel bzw. an ein klinisches Instrument formuliert, um es

- in dem sehr speziellen klinischen Umfeld der Sprechstörungsdiagnostik,
- unter Berücksichtigung der Anwender-Zielgruppe (Logopäden, Sprachheilpädagogen, Linguisten, Mediziner),
- unter dem Aspekt der zeitlichen Verfügbarkeit bzw. der tolerablen zeitlichen Bindung von Personal und/oder Patient

tatsächlich einer späteren routinemäßigen Nutzung durch klinische Praktiker zuführen zu können. Ausgehend von einer klinisch-diagnostischen Bedarfsanalyse erfolgt im Anschluß eine konkrete Spezifizierung bzw. die Definition von Zielen, die im Rahmen der Entwicklung verfolgt werden sollen. Bedingt durch die langjährige Erfahrung und die angesammelte Fachexpertise der EKN bzw. der Abteilung für Klinische Neuropsychologie München-Bogenhausen konnten diese Anforderungen und die abgeleiteten Ziele an das MODIAS-System bereits zu Beginn der Entwicklung zügig und relativ detailliert erarbeitet werden.

Der folgenden Abschnitt beschreibt die angestrebten Untersuchungsverfahren, sowie die Systematik und den Aufbau der entsprechenden Untersuchungsprotokolle.
Im letzten Teil des Kapitels werden die technischen Rahmenbedingungen definiert mit Konkretisierungen bezüglich Zielplattform, Hardware-/Softwarekonzept und gewählter Entwicklungsumgebung.

3.1. Analyse des Bedarfs und Definition der Ziele

3.1.1. Praxis der klinischen Sprechstörungsdiagnostik

Die auditive Befundung hat in der klinischen Diagnostik aphasischer und sprechapraktischer Störungen der Sprachproduktion bisher noch immer eine zentrale Bedeutung. Akustisch-phonetische Untersuchungsverfahren tragen hier aber zur Differenzierung und Objektivierung des klinischen Bildes in einem ganz erheblichen Maße bei. Neuere Entwicklungen der letzten Jahre, vor allem die rasante Entwicklung der Rechnerleistungen resultierten in einer breiten Verfügbarkeit auch PC-gestützter Systeme zu akzeptablem Preis.
Viele Sprachtherapeuten schrecken aber vor dem routinemäßigen Einsatz solcher Verfahren im klinischen Alltag zurück. Die meisten Programme und Systeme liefern eine Vielzahl berechenbarer Parameter z.B. für Stimmqualitätsanalysen, allerdings sind diese oft nur schwer überschaubar und sie verfügen teilweise nur über relativ geringe diagnostische Relevanz. Zu guter letzt decken viele Systeme nicht alle Aspekte des klinischen Routinealltages ab, die bei einer auf akustischen Messungen gestützten Standarduntersuchung relevant wären, wie z.B. Stimuluspräsentation, Sprachaufzeichnung und -wiedergabe, Signalanalyse, Artefakterkennung und Ergebnispräsentation.
Das hier vorgestellte System MODIAS ist ein Versuch, die aus den Erfahrungen abgeleiteten Erfordernisse weitgehend abzudecken, verbunden mit neuesten Methoden und Verfahren der Digitalen Signalverarbeitung. Die Kombination medizinischer Expertise mit ingenieurwissenschaftlicher Methodik läßt hier Synergieeffekte erwarten, die unter konsequenter Einbeziehung therapeutischer Erfahrungswerte zu einer praktikablen und im klinischen Umfeld anwendbaren Systemlösung führen sollte.

3.1.2. Diagnostische und technische Zielsetzung

Das System MODIAS zielt auf die Erarbeitung akustisch-phonetischer Standards zur Verbesserung der klinischen Diagnostik neurogener Sprechstörungen, also von Störungen der zentralmotorischen Programmierungs- und Steuerungsprozesse des Sprechbewegungsablaufes. Die konkret gestellten Anforderungen und die zu erreichenden Ziele wurden vor Beginn der Entwicklung wie folgt definiert:

- Das Standard-Untersuchungsprotokoll soll umfassend sein, aber dennoch ökonomisch und einfach in der Durchführung.
- Die wichtigen Leistungsbereiche Tempo, Rhythmus, Sprechflüssigkeit, Stimme, Artikulation und Diadochokinese sollen geprüft werden.
- Quantitative Leistungsmaße sollen durch akustische Parameter abgeleitet werden, aber weitgehend ohne zeitraubende Segmentierungsarbeit und bei einfacher Handhabung.
- Die Berechnung und Verarbeitung der Daten kann Offline erfolgen, damit sollte eine Patientensitzung und eine Auswertungssitzung jeweils getrennt durchführbar sein.
- Die Befunddarstellung soll übersichtlich sein und dabei dem Untersucher Grafiken, Meßwerte und Vergleichswerte zur Verfügung stellen.
- Die Gesamtkosten für Hard- und Software sollen so gering wie möglich gehalten werden, um eine breite klinische Nutzung auch bei geringem Budget zu ermöglichen.

3.2. Untersuchungsverfahren und Diagnostisches Konzept

Die beschriebenen Anforderungen wurden bei der Realisierung von MODIAS in ein Untersuchungsprotokoll mit 4 unabhängigen Modulen umgesetzt.
Organisatorisch erfolgt die Patientenuntersuchung ausschließlich durch Analyse des über Mikrophon aufgezeichneten Sprachsignales, also ohne störende Meßaufnehmer im Mund- bzw. Rachenraum oder im Kehlkopfbereich des Patienten (Abbildung 3.1).

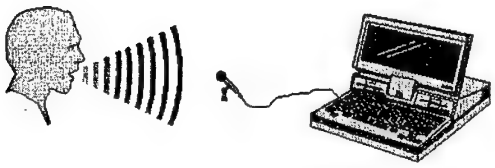

Abbildung 3.1.: Die Patientenuntersuchung erfolgt innerhalb des MODIAS-Systems ausschließlich an dem über Mikrophon und Soundkarte aufgezeichneten Sprachsignal, somit ohne beeinflussende Sensorik in Mund- bzw. Rachenraum oder im Kehlkopfbereich.

Die Aufzeichnungen werden durch den Untersucher durchgeführt, alle nachgeordneten Berechnungen und Analysen erfolgen Offline ohne zeitliche Bindung des Patienten, ggf. auch zu einem späteren Zeitpunkt.

Das Modul STIMME UND VOKALARTIKULATION zielt mit seinen diagnostischen Fragestellungen sowohl auf die Diagnostik von Stimmstörungen als auch auf die Prüfung des Artikulationsraumes und der Artikulation der Vokale. Dieses Modul ist primär für dysarthrische Patienten mit ihren spezifischen Störungsprofilen konzipiert.
Für den Bereich der Stimmstörungen werden Stimmstabilität und -qualität analysiert mit resultierenden Aussagen über Stimmlage, Tonhöhenschwankungen, Rauhigkeits- oder Behauchtheitsphänomene.
Für den Bereich der Vokalartikulation werden die charakteristischen Resonanzfrequenzen des Artikulationstraktes und deren zeitlicher Verlauf geprüft mit abgeleiteten Parametern bezüglich der Fähigkeit zur artikulatorischen Kontrastierung und zur korrekten Konfigurierung der Artikulatoren (Zungenposition, Kieferöffnung und Lippenrundung).
Das Untersuchungsprotokoll sieht in diesem Modul die Aufzeichnung isolierter, gehaltener Vokale (*/i/*, */y/*, */u/* und */a/*) vor, mit jeweils 3 Durchgängen.

Das Modul SATZPRODUKTION widmet sich dem Störungsprofil einer gestörten Satzproduktion. Ein veränderter Redefluß gehört zu einem der häufigsten Symptome neurogener Sprechstörungen. Er kann sowohl erhöht als auch verlangsamt sein, letzteres wird in der Praxis jedoch wesentlich häufiger beobachtet ([Ack92]). Ein verlangsamter Redefluß kann vielerlei Ursachen haben, unter anderem eine zeitliche Dehnung der

Vokale, häufige oder verlängerte Pausen, Wiederholungen, Korrekturversuche usw.
Das beschriebene Modul zielt auf die klinische Diagnostik vorwiegend aphasischer und sprechapraktischer Störungsprofile und bietet hierfür eine Methode zur akustischen Analyse des Zeitmusters der Satzproduktion bei gleichzeitiger auditiver Analyse nach phonetischen und phonematischen Fehlern. Es eignet sich daher zur Differenzierung unterschiedlicher Störungsmuster und zur Analyse der Faktoren, die das Auftreten von Unflüssigkeiten beeinflußen.
Der Patient hat hier die Aufgabe, 24 ausgewählte Sätze nachzusprechen, die sich nur in einem Zielwort unterschiedlicher Komplexität unterscheiden. Der Trägersatz ist stets identisch (*Ute kann die ... bekommen*), die eingebetteten Zielwörter bestehen zu je einem Drittel aus einfachen Zweisilblern der Konsonant-Vokal-Struktur (z.B. *Kette*), aus Zweisilblern mit Konsonant-Clustern (z.B. *Pflanze*) und aus gemischten Dreisilblern (z.B. *Trompete*).
Für die Aufgabenstellung der Therapiekontrolle, die in der Regel mit mehreren, zeitlich dicht aufeinanderfolgenden Untersuchungen am gleichen Patienten einhergeht, wurde zur Vermeidung von Lerneffekten ein alternatives Satzensemble implementiert, welches zusätzlich auch die Trägerphrase variiert (bei einer jedoch unveränderlichen zeitlichen bzw. phonetischen Struktur und gleicher Zielwortmenge).

Das Modul SCHNELLE SILBENWIEDERHOLUNGEN zielt auf die Untersuchung der Fähigkeit zu schnellen Silbenwiederholungen (Artikulatorische Diadochokinese). Die meisten Dysarthrieformen sind durch eine Verlangsamung gekennzeichnet, parallel zum Ausmaß der Artikulationsstörung. Wechselndes Tempo, Abnahme des Tempos, dysrhythmische Lautstärkeschwankungen sowie wechselnde Pausen charakterisieren die ataktische Dysarthrie ([Poe89]).
Der Patient wird in diesem Modul aufgefordert, so schnell wie möglich fortlaufend die Silben /ba/,/da/,/ga/,/na/ sowie die Kombinationen /bada/ und /dana/ zu sprechen. Silbengrenze und Silbenkerne werden bezüglich zeitlicher Lage und Amplitude vermessen und daraus Parameter abgeleitet, die Tempo, Konstanz und Regularität beschreiben.

Das Modul FRIKATIVARTIKULATION widmet sich primär dem Störungsprofil unpräziser Konsonantartikulation, welche in schweren Fällen zu erheblichen Verständlichkeitsproblemen führen kann. Die meisten dysarthrischen Syndrome sind durch solche artikulatorische Defekte charakterisiert. Geprüft wird in diesem Modul sowohl die Fähig-

keit zu einer kontrastreichen Artikulation der 3 stimmlosen Zischlaute /s/, /ʃ/ und /x/ gegenüber einem jeweils vorangehenden Vokal (*Kontrastierung Frikativ/Vokal*) als auch zu einer ausreichenden Unterscheidung dieser 3 Frikative (*Frikativ-Differenzierung*).
Hierzu wurde ein geeignetes Wortmaterial zusammengestellt in Form der beiden Einzelwort-Tripel *lassen, Laschen, lachen* und *Massen, Maschen, machen*. Das Untersuchungsprotokoll sieht die Aufzeichnung der beiden beschriebenen Wort-Ensembles in jeweils 6 Durchgängen vor.
Die Prüfung der Kontrastierung Frikativ/Vokal erfolgt hierbei durch die Überwachung des Klassifikationsverhaltens eines speziell entworfenen Mustererkenners, der auf die akustischen Eigenschaften dieser beiden Lautklassen anhand einer Stichprobe von Normsprechern trainiert wurde.
Die Prüfung der Frikativ-Differenzierung erfolgt in ähnlicher Weise, hier jedoch ausschließlich anhand der Klassifikationsergebnisse eines zweiten Mustererkenners. Er wurde speziell auf die akustischen Eigenschaften der 3 betrachteten Frikativklassen anhand der gleichen Stichprobe trainiert. Zusätzlich ist im Untersuchungsprotokoll eine auditive Komponente berücksicht, die eine Klassifikation der Frikative auf Basis perzeptiver Urteile gestattet.
Die beiden Entscheidungsinstanzen *Mensch* und *Maschine* können in ihrem Urteilsverhalten verglichen werden.

3.3. Technisches Konzept

3.3.1. Gewähltes Zielsystem

Abgeleitet aus den Rahmenbedingungen möglichst geringer Systemkosten einerseits, und der vergleichsweise niedrigen Geschwindigkeitsanforderungen (Offline-Verarbeitung) andererseits, wurde für die Realisierung des MODIAS-Systems das folgende Zielsystem gewählt:

- IBM-kompatibler Standard-PC,
- Betriebssystem MS WINDOWS in der Version 3.1 oder höher,
- Arbeitsspeicher mindestens 32 Megabyte,

- Qualitativ hochwertige, IBM-kompatible Soundkarte (Abtastrate 22050 Hz, Quantisierungsauflöung 16 Bit),
- Dynamisches Studiomikrofon mit 600Ω Eingangsimpedanz.

Die Wahl für einen Standard-PC resultiert vor allem aus der Tatsache, daß diese Hardwareplattformen heute praktisch überall vorhanden sind. Bei der Neueinrichtung eines MODIAS-Untersuchungsplatzes entstehen somit in der Regel keine diesbezüglichen Zusatzkosten. Zu Beginn der MODIAS-Realisierung wurde die zukunftsweisende Entscheidung getroffen, sich sowohl in der Entwicklungsphase als auch in der späteren Phase der klinischen Nutzung auf die kommerziellen Softwareprodukte

- MATLAB (Fa. MathWorks Inc., USA) mit der optionalen Ergänzung
- SIGNAL PROCESSING TOOLBOX.

abzustützen, um die Entwicklungszeiten minimieren und eine frühzeitige klinische Erprobung einzelner, bereits fertiggestellter Module zu ermöglichen.

Anmerkung: Die MODIAS-Software besteht im Prinzip, wie der Großteil von MATLAB selbst, aus einer Sammlung spezieller ASCII-Source-Codes bzw. *m-Files*, die vom MATLAB-Kernel jeweils erst zur Laufzeit interpretiert werden. Die m-Files können auf jede beliebige Hardware (entsprechende MATLAB-Installation vorausgesetzt) eingespielt werden, damit ist die Programmierung völlig hardwareunabhängig.
Schon 1994 war erkennbar, daß die Philosophie der Firma MathWorks einerseits, und die weltweite Verbreitung und vorwiegend wissenschaftliche Nutzung des MATLAB-Produkts andererseits, auf eine spätere Kompilierbarkeit vorhandener m-Files zur Erzeugung von Stand-Alone-Applikationen zielen würde. Seit Juli 1999 ist es nun tatsächlich möglich, aus m-Files (bzw. komplex verknüpften m-File-Sammlung wie z.B. MODIAS) eigenständig lauffähige Programme für verschiedene Hardware-Plattformen zu generieren. Die Anzahl der Kompilationen ist hierbei nicht beschränkt, die Nutzung unterliegt keinen lizenzrechtlichen Einschränkungen. Zum Zeitpunkt der Erstellung der vorliegenden Arbeit lagen hierzu noch keine Erfahrungen vor, gem. Aussage der deutschen Vertriebsfirma ist die Erfolgsaussicht einer fehlerfreien MODIAS-Kompilation jedoch optimistisch

zu bewerten. Damit entfiele für die MODIAS-Anwendung die beschriebene Abstützung auf MATLAB, verbunden mit einer nicht unerheblichen finanziellen Entlastung für den klinischen Anwender bzw. dessen Betriebsorganisation.

3.3.2. Verwendete Entwicklungsumgebung

Als Entwicklungsumgebung diente während der gesamten Laufzeit ein handelsüblicher PC (INTEL-Pentium-90-Prozessor) mit 32 MB Arbeitsspeicher. Zur Entwicklung von Algorithmen und Bedienoberfläche des MODIAS-Systems wurde der mathematische Interpreter MATLAB (Version 4.2c) der Firma MathWorks mit der optionalen SIGNAL PROCESSING TOOLBOX gewählt, der sich in der Digitalen Signalverarbeitung mittlerweile als weltweiter Standard etabliert hat.
MATLAB gestattet einen einfachen Zugriff auf sämtliche mitgelieferte Standardroutinen sowie deren Einbindung in benutzerspezifische Projekte. Damit entfallen nichtkreative Programmierzeiten für Standardfunktionalitäten wie z.B. grafische Ausgaben, akustische Wiedergabe von Sprachdaten, Papierausdruck, Dateiablage oder auch für die bereits hinlänglich bekannten Algorithmen wie z.B. zur Fast-Fourier-Transformation oder auch zur digitalen Filterung. MATLAB bietet somit eine optimale Umgebung für die Entwicklung und die Tests neuentwickelter Signalverarbeitungsalgorithmen, besonders aufgrund des Interpreterkonzeptes und des damit verbundenen Wegfalls von Compilierzeiten.
Das Aufzeichnungssystem bestand aus einer qualitativ sehr hochwertigen Soundkarte, Typ TROPEZ der Firma TURTLE BEACH, sowie einem dynamischen Studiomikofon mit einer Eingangsimpedanz von 600Ω, Typ DM130 der Firma VIVANCO .

4. Realisierung des Moduls STIMME UND VOKALARTIKULATION

Das diagnostische und technische Konzept zu dem Modul STIMME UND VOKALARTIKULATION wurde bereits im vorhergehenden Kapitel aufgezeigt. Nach einer kurzen Erläuterung der relevanten diagnostischen Fragestellungen widmet sich dieses Kapitel der konkreten Realisierung innerhalb des MODIAS-Systems. Der interdisziplinäre Charakter der Aufgabenstellung läßt hierbei eine Zweiteilung sinnvoll erscheinen in die

- klinisch/diagnostisch relevanten Anteile, wie z.B. Untersuchungsprotokoll, gewähltes Untersuchungsmaterial, Ablauf einer Untersuchungssitzung und Auswertung bzw. Darstellung der Ergebnisse, sowie in
- technisch relevante Anteile, wie z.B. angewandte bzw. entwickelte technische Methoden, Verfahren und Algorithmen, die zu der ingenieurwissenschaftlichen Lösung der medizinischen Problemstellung geführt haben.

Bedingt durch die grundsätzlich schwierige Übertragbarkeit akustischer Stimmparameter auf den perzeptiven Höreindruck wurde hier auf die Validität besonderes Augenmerk gelegt. Dabei wurde der methodische Ansatz des Experiments (Isolierte Vokale) zunächst anhand einer Pilotstudie überprüft. Die Sensitivität der Stimmqualitätsmaße wurde anschließend an synthetischem Sprachmaterial validiert. Der Entwurf und die Realisierung des zur Erzeugung des synthetischen Sprachmaterials erforderlichen Vocoders sind ergänzend dargestellt.
Im letzten Teil des Kapitels werden die Ergebnisse einer typischen klinischen Anwendung vorgestellt, bei der die Stimmparameter von Patienten mit neurogenen Sprechstörungen

mit den entsprechenden Parametern sprechgesunder Probanden verglichen wurden. Damit konnte ein klinischer Nachweis der Sensitivität des MODIAS-Moduls erbracht werden. Ergänzend werden die Ergebnisse einer klinischen Anwendung im Rahmen einer Kooperation mit dem Klinikum Rechts der Isar der Technischen Universität München vorgestellt, deren Ergebnisse bereits vielfach publiziert worden sind (z.B. in [Bre98]). Es handelt sich hierbei um die bisher umfassendste Untersuchung an postoperierten Patienten mit Lippen-Kiefer-Gaumenspalte, also einer Patientengruppe mit rein organisch bedingten Sprechstörungen ohne neurologische Schäden.

4.1. Diagnostische Fragestellungen

Die aus den realisierten Vokalen */i/*, */y/*, */u/* und */a/* abgeleiteten Formanten (charakteristische Resonanzfrequenzen des menschlichen Vokaltrakts, siehe Seite 16 ff. und 33 ff.) stellen ein Maß für die Bewegungsauslenkung der Artikulatoren dar. Das gebräuchlichste Maß für die Beurteilung der Vokalqualität ist die Lage der beiden ersten Formanten ([Alf82]). Hierzu leitet MODIAS neben den zeitlichen Verläufen auch die mittleren Werte für die Formanten F1 und F2 ab.

Bei einer Interpretation dieser Daten geht man davon aus, daß der erste Formant mit dem Öffnungsgrad der Vokale korreliert ([Tra88]). Der Befund eines erhöhten ersten Formanten weist in diesem Sinne auf eine zu geringe Anhebung der Zunge bei der Bildung der hohen Vokale */i/*, */y/* und */u/*, also auf eine Tendenz zur Zentralisierung, hin ([Zie83a],[Zie83b]). Entsprechend kann eine Absenkung des ersten Formanten beim Vokal */a/* durch eine verengte artikulatorische Konfiguration, etwa infolge eines zu geringen Kieferöffnungswinkels, interpretiert werden.
Die Lage des zweiten Formanten wird einerseits durch die Rundung und Protrusion der Lippen, andererseits durch die horizontale Zungenposition beeinflußt. Durch Vergleich der zweiten Formanten für die Vokale */i/* und */y/* erhält man ein Maß für die Rundung der Lippen. Die F2-Differenz von */y/* und */u/* stellt ein Maß für die Kontrastierung dieser beiden Vokale durch Vor- (*/y/*) bzw. Rückverlagerung (*/u/*) der Zunge dar.

Fast alle neurologischen Erkrankungen, welche die generelle motorische Steuerung beeinträchtigen, wirken sich auch auf die Kehlkopfsteuerung aus mit der Folge von Stö-

rungen der Stimmbildung (Stimmstörungen). Üblicherweise ist die neurogene Dysphonie nur Teil einer allgemeineren dysarthrischen Sprechstörung, die auch die motorische Steuerung der Atmungs- und der supralaryngealen Muskulatur während des Sprechvorganges betreffen. Die wahrnehmbaren Stimmstörungen dieser Patienten reflektieren sowohl die direkte Beeinträchtigung des laryngealen motorischen Systems als auch die kompensatorische Reaktionen auf respiratorische oder artikulatorische Defekte ([Zie99],[Merk99a]).

Die klinische Diagnostik neurogener Stimmstörungen basiert in erster Linie auf perzeptiven Bewertungen der Stimmqualität ([Dar75]). Dieses Verfahren ist jedoch naturgemäß subjektiv und in den meisten Fällen nicht ausreichend zuverlässig, um spontane oder therapiebedingte Fortschritte zu verfolgen. Objektivere Verfahren basieren hier auf Methoden der Digitalen Signalverarbeitung.
Diese (akustischen) Verfahren zur Beurteilung von Stimmstörungen sind in der klinischen Diagnostik und der Therapieüberwachung noch nicht etabliert. Von der hohen Korrelation zwischen den zahlreichen akustischen Parametern wie, z.B. *Jitter*, *Shimmer*, oder *Harmonics-to-Noise-Ratio* und den perzeptiv bewerteten Stimmqualitäten, wie *Rauhigkeit*, *Heiserkeit*, *Behauchtheit* oder *Gepreßtheit*, wurde in zahlreichen Publikationen berichtet (z.B. [Bak87], [Chi91], [Esk90], [Mil87]).

4.2. Aufbau und Gestaltung

4.2.1. Untersuchungsprotokoll und -material

Der Patient hat in diesem Modul die Aufgabe, die Kardinalvokale */i/*, */y/*, */u/* und */a/* über einen Zeitraum von mindestens 3 Sekunden in einer aus seiner Sicht normalen Intensität und Sprechstimmlage zu vokalisieren. Der Untersuchungsgang sieht 3 Durchgänge vor, also entstehen insgesamt 12 Aufzeichnungen bzw. 3 Aufzeichnungen pro Vokal.

4.2.2. Untersuchungsablauf und Bedienung

Nach Aufruf von MODIAS erscheint die Startoberfläche gem Abbildung 4.1 mit dem entsprechenden Startbutton für den Untersuchungsteil STIMME UND VOKALARTIKULATION.

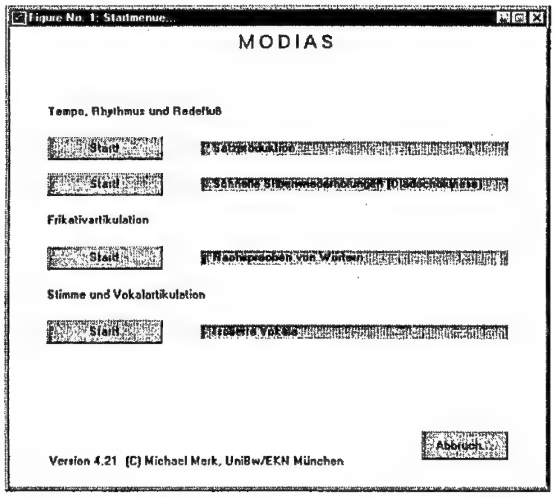

Abbildung 4.1.: Startoberfläche des MODIAS-Systems. Das Modul STIMME UND VOKALARTIKULATION wird mit dem zugehörigen Button START! aufgerufen.

Nach Betätigung des Buttons START! gelangt man unmittelbar zur Organisationsoberfläche des Moduls STIMME UND VOKALARTIKULATION gem. Abbildung 4.2. An dieser Stelle wird der Untersucher aufgefordert, sämtliche relevanten Daten zum Patienten und zur aktuellen Untersuchungssitzung in die vorgesehenen Felder einzugeben. Diese Organisationsdaten bestehen aus:

- PATIENTENKÜRZEL: Maximal 7-stellige Kennung zur eindeutigen Identifizierung des Patienten. Umlaute oder Leerzeichen werden automatisch ersetzt, da syntaktisch vom System nicht vorgesehen.
- GESCHLECHT: Geschlecht des Patienten. Auswahl des entsprechenden Buttons für weiblich bzw. männlich.
- ALTER: Lebensalter des Patienten in Jahren zum Zeitpunkt der Aufzeichnung bzw. Untersuchungssitzung.

- DATUM: Feld wird automatisch mit dem aktuellen Datum gem. eingestellter Systemzeit des Rechners vorbefüllt, kann aber editiert werden.
- SESSION NR: Laufende Nummer der Sitzung an dem im Feld DATUM eingetragenen Tag.

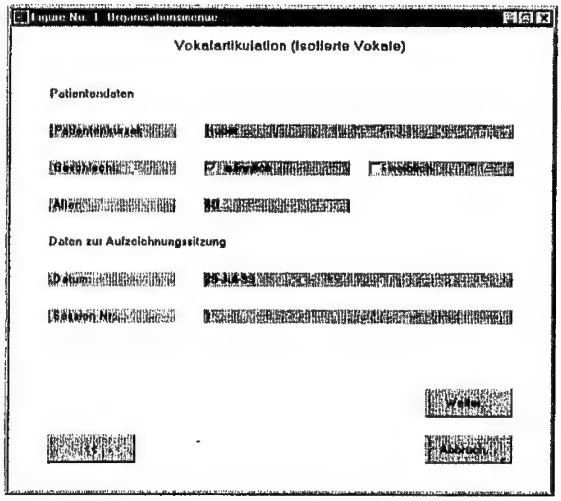

Abbildung 4.2.: Organisationsoberfläche des Moduls STIMME UND VOKALARTIKULATION zur Eingabe sämtlicher Organisationsdaten bezüglich des Patienten und der Untersuchungssitzung.

Der Button < < führt unmittelbar zum Rücksprung in die Startoberfläche des MODIAS-Systems gem. Abbildung 4.1.
Nach Betätigung des Buttons WEITER wird systemintern aus den eingegebenen Daten der Felder PATIENTENKÜRZEL, DATUM und SESSION NR der Ablagepfad für alle im Laufe der Untersuchungssitzung generierten Dateien abgeleitet. Sämtliche eingegebene Organisationsdaten werden dabei permanent gespeichert, d.h., bei Aufruf einer früheren Sitzung reicht zur korrekten Identifizierung die Eingabe von Patientenkürzel, Datum und Sitzungsnummer aus, die restlichen Informationen über Alter und Geschlecht des Patienten werden aus den gespeicherten Daten rekrutiert und in den entsprechenden Feldern

eingetragen (In diesem Fall Feld SESSION NR mit <Return> verlassen, danach erfolgt die beschriebene Datenrekonstruktion).
Der Untersucher wird unmittelbar zur Steuerungsoberfläche gem. Abbildung 4.3 weitergeleitet, die eine Festlegung bezüglich des weiteren Ablaufs der Sitzung fordert bzw. ermöglicht.

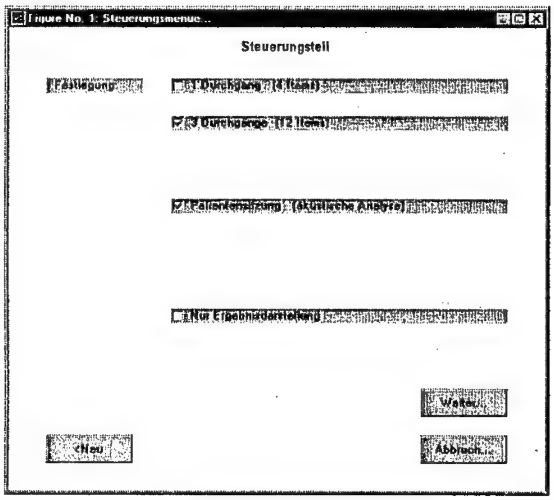

Abbildung 4.3.: Steuerungsoberfläche des Moduls STIMME UND VOKALARTIKULATION. An dieser Stelle wird vom Untersucher der Umfang des Untersuchungsprotokolls festgelegt bzw. direkt zur Auswertung (bei bereits vorliegenden Ergebnissen) navigiert.

Im oberen Teil der Steuerungsoberfläche kann der Untersucher auswählen, ob das

- komplette Standardprotokoll mit 12 Aufzeichnungen und damit je 4 pro Vokal oder aber das
- verkürzte Untersuchungsprotokoll mit 4 Aufzeichnungen und damit je 1 pro Vokal

genutzt werden soll.

Im unteren Teil der Steuerungsoberfläche wird der weitere Ablauf bzw. die Zielrichtung der Sitzung festgelegt:

- PATIENTENSITZUNG: Erstellung neuer Aufzeichnungen mit anschließender Analyse und Auswertung;
- ERGEBNISDARSTELLUNG: Sichtung bereits vorhandener Ergebnisse (ohne Aufzeichnung).

Bei Auswahl der Patientensitzung gelangt man nach Betätigung des Buttons WEITER zur ersten der bis zu 3 aufeinanderfolgenden Aufzeichnungsoberflächen gem. Abbildung 4.4 (bei Auswahl der Ergebnisdarstellung direkt zur Auswertungsoberfläche gem. Abbildung 4.8 auf Seite 56).

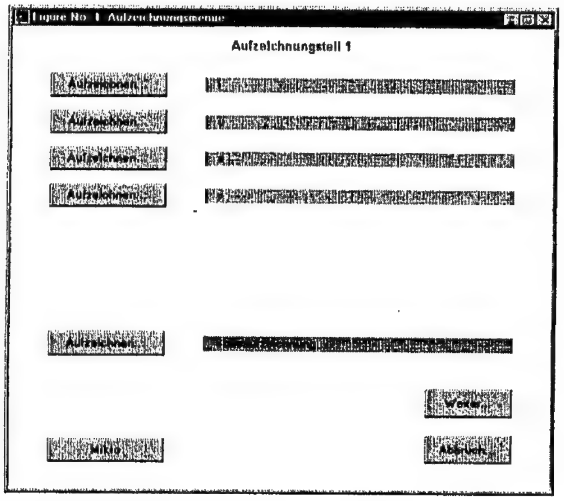

Abbildung 4.4.: Eine der bis zu 3 aufeinanderfolgenden Aufzeichnungsoberflächen des Moduls STIMME UND VOKALARTIKULATION. Die Reihenfolge der Vokale ist randomisiert. Die Aufzeichnung wird jeweils durch den Button AUFZEICHNEN gestartet.

Die Darstellung der Vokale erfolgt randomisiert, die Aufzeichnung jedes Vokals wird

durch Betätigung des entsprechenden Buttons AUFNEHMEN gestartet. Da MATLAB in der verwendeten Version noch keinen direkten Zugriff (bei der Aufnahme) auf die Soundkarte ermöglicht, erscheint das standardisierte MS-WINDOWS-Tool des SOUND-RECORDERS gem. Abbildung 4.5.

Abbildung 4.5.: Bedienoberfläche des verwendeten Tools zur Steuerung der Aufzeichnung. Aus Kompatibilitätsgründen wird in allen MODIAS-Modulen einheitlich der mit dem Betriebssystem MS-WINDOWS ausgelieferte *Soundrecorder* genutzt.

Die Aufzeichnung wird über die Start- und Stopptaste gesteuert, eine unmittelbare Wiedergabe zur Kontrolle ist möglich. Liegt bereits eine Aufzeichnung vor, wird innerhalb der Aufzeichnungsoberfläche der Button AUFNEHMEN automatisch durch den Button ANHÖREN ersetzt und damit eine entsprechende Wiedergabemöglichkeit bereitgestellt. Bei der Aufzeichnung wird jeder Vokal vom Patienten in normaler Sprechstimmlage bei normaler Intensität über einen Zeitraum von mindestens 3 sec anhaltend gesprochen. Soll die Aufzeichnung gespeichert werden, genügt das Schließen des Soundrecorders und die Speicherung wie vorgeschlagen. Der Dateiname ist durch MODIAS bereits voreingestellt, ebenso die Systemeinstellungen bezüglich Samplingrate (22050 Hz) und Auflösung (16Bit).
Die im Untersuchungsprotokoll vorgesehene Stilleaufzeichnung sieht eine Aufzeichnung ohne Aufsprache bzw. ohne jegliche Art von (nicht systemimmanenten) Störgeräuschen mit einer Dauer von ca. 5 sec vor. Eine automatisch durchgeführte statistische Analyse des systemspezifischen Rauschens des Aufzeichnungssystems Soundkarte/Rechner ermöglicht eine wesentlich robustere Festlegung für die spätere Festlegung günstiger Signalabschnitte.
Wie bereits dargestellt, besteht jede Aufzeichnung aus einem mindestens 3 sec möglichst konstant realisierten Vokal. Bei Patienten mit neurogenen Sprechstörungen kann diese gestellte Aufgabe bereits zu Problemen führen. Die klinische Praxis zeigt z.B. das Auftreten

von Unterbrechungen, Abbrüchen, Räuspern oder auch unerwünschten Geräuschartefakten. Dem Untersucher soll hier ein Instrument an die Hand gegeben werden, mit dem er die nach seiner Einschätzung verwertbaren und im Sinne der zu diagnostizierenden Stimmstörung repräsentativen Anteile der Aufzeichnung in die weitere Analyse mit einbeziehen und andere Anteile verwerfen kann.
Hierzu wurde in dem Modul STIMME UND VOKALARTIKULATION ein spezieller Segmentierungseditor implementiert, der einen mit der subjektiven Lautheitsempfindung des menschlichen Gehörs korrespondierenden Kurvenverlauf (Lautheitskontur) darstellt und damit eine geeignete Oberfläche für die Extraktion verwertbarer Signalsegmente bereitstellt.
Die erforderlichen Berechnungen der Lautheitskonturen sowie der vom System aus akustischer Sicht vorgeschlagenen Segmentgrenzen werden automatisch beim Übergang von den Aufzeichnungsoberflächen zur anschließenden Segmentierungsoberfläche gem. Abbildung 4.6 für jeden der realisierten Vokale getrennt berechnet.
Die Segmentierungsoberfläche gestattet für jede durchgeführte Aufzeichnung (erkennbar an einer gelben Markierung am rechten Rand der jeweiligen Textfelder) über den Button SEGMENTIEREN den Aufruf des Segmentierungseditors gem. Abbildung 4.7.
Im Segmentierungseditor wird die berechnete Lautheitskontur des jeweiligen Vokals grafisch darstellt. Vom System wird dabei bereits ein zusammenhängendes Signalsegment vorgeschlagen und entsprechend markiert, das

- aus akustischer Sicht günstige Signaleigenschaften aufweist bzw. mit geringer Wahrscheinlichkeit störende Artefakte enthält und
- eine zeitliche Dauer von bis zu 2 sec besitzt.

Die beiden rechts oben eingeblendeten Informationsfelder zeigen den aktuellen Stand der Segmentierung an und sind wie folgt zu interpretieren:

- ANGESTREBT: Idealerweise zu erreichende Dauer des segmentierten Signalsegments von 2 Sekunden. Die Anzeige ist statisch und unabhängig von der aktuellen Segmentierung.
- ERKANNT: Tatsächliche Dauer des segmentierten Signalsegments in Sekunden, die sich aus dem Abstand der beiden eingeblendeten senkrechten Linien ergibt. Die Anzeige ist dynamisch und hängt von der aktuellen Segmentierung ab.

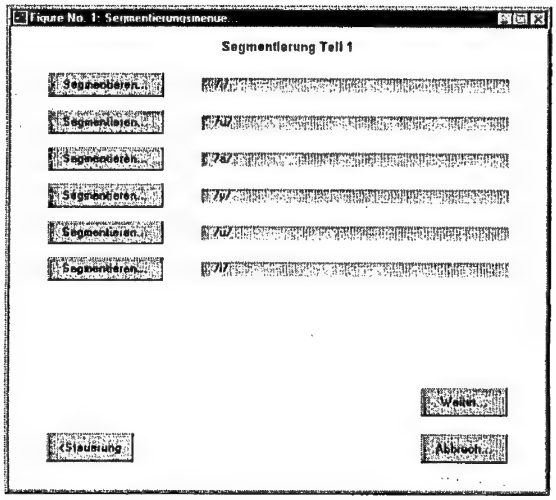

Abbildung 4.6.: Segmentierungsoberfläche des Moduls STIMME UND VOKALARTIKULATION. Durch Betätigung des Buttons SEGMENTIEREN wird die halbautomatische Segmentierung eines günstigen Signalabschnittes eingeleitet.

Müssen die vorgeschlagenen Segmentgrenzen bezüglich ihrer zeitlichen Lage korrigiert werden, stehen dem Untersucher mehrere Interaktionsmöglichkeiten zur Verfügung. Unterstützend wird hierzu im unteren Teil des Bildschirms das zeitsynchrone Oszillogramm eingeblendet.

Das Menü MANIPULATION dient der unmittelbaren, manuellen Einflußnahme auf die bisher automatisch erfolgte Segmentierung. Es bietet die folgende Auswahl:

- LINKE bzw. RECHTE GRENZE VERSCHIEBEN: Erlaubt die manuelle Verschiebung der Segmentgrenzen. Der Segmentkern wird hierfür mit der linken Mousetaste selektiert, danach wird der neue Ort der linken bzw. rechten Segmentgrenze mit gleicher Mousetaste vorgegeben.

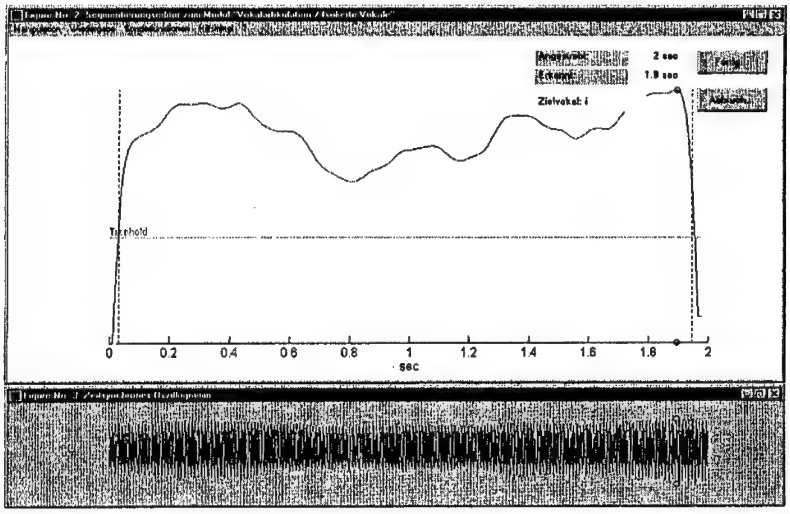

Abbildung 4.7.: Bedienoberfläche des Segmentierungseditors des Moduls STIMME UND VOKALARTIKULATION. Hier erfolgt die manuelle Einflußnahme des Untersuchers auf die vom System vorgeschlagene Segmentierung eines günstigen Signalabschnitts.

- SEGMENT VERSCHIEBEN: Erlaubt die manuelle Verschiebung des gesamten Segments. Der Segmentkern wird hierfür mit der linken Mousetaste selektiert, danach wird der neue Ort des Segmentkerns mit gleicher Mousetaste vorgegeben.

Das Menü WIEDERGABE gestattet dem Untersucher die perzeptive Kontrolle der bisherigen Segmentierung. Die Auswahl erlaubt die Wiedergabe der gesamten Aufzeichnung oder die Wiedergabe des markierten Segments.

Das Menü SONDERFUNKTIONEN bietet dem Untersucher die Möglichkeit, eine spezifische Mindestlautheit für die Detektion eines Segments direkt in der Grafik über die linke Mousetaste vorzugeben (THRESHOLD FESTLEGEN).

Das Menü HILFSMITTEL ermöglicht eine benutzerdefinierte Zoomfunktion be-

züglich der Zeitachse bei fester Skalierung der Lautheitsachse. Die Festlegung des darzustellenden Bereiches erfolgt durch Aufziehen eines Zoomfensters bei gedrückter linker Mousetaste. Die drei weiteren Auswahlmöglichkeiten dieses Menüs steuern das Ein- bzw. Ausblenden des Oszillogrammes sowie die Anpassung der Oszillogramm-Darstellung auf bereits erfolgte Manipulationen (linke bzw. rechte Silbengrenze verschoben) oder geänderte Skalierungen (Zeitachse gezoomt).

Das letzte Menü ? bietet eine Hilfefunktion mit einer ausführlichen Beschreibung zur Bedienung des Segmentierungseditors.
Nach Verlassen des Segmentierungseditors mit dem Button OK werden die jetzt überflüssigen Signalabschnitte vor der linken und nach der rechten Segmentgrenze entfernt. Die Aufzeichnung wird damit zum Teil erheblich gekürzt, was sich in einer deutlichen Ersparnis im Speicherbedarf auswirkt.
Sind alle Aufzeichnungen innerhalb der Segmentierungsoberfläche bearbeitet, erfolgt der Übergang zur Auswertung.

4.2.3. Auswertung, Ergebnisdarstellung und Datenexport

Die Auswertungsoberfläche gem. Abbildung 4.8 gestattet die getrennte Auswertung der Aufzeichnungen nach diagnostischen Schwerpunkten:

- ARTIKULATIONSANALYSE: Prüfung der artikulationsspezifischen Signalparameter.
- STIMMANALYSE: Berechnung der für Stimmqualität und Stimmstabilität spezifischen Signalparameter.

4.2.3.1. Auswertung der Vokalartikulation

Die ersten 4 Auswerteblätter gem. Abbildung 4.9 zeigen jeweils auf der linken Seite die zeitlichen Verläufe der Formantfrequenzen F1 und F2 sowie deren Mediane. Jedem der 4 Vokale ist ein eigenes Auswertungsblatt zugeordnet, es enthält dabei sämtliche Realisierungen der Sitzung.

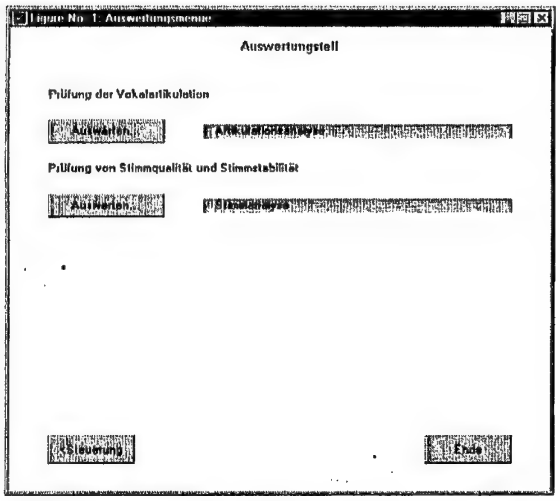

Abbildung 4.8.: Auswertungsoberfläche des Moduls STIMME UND VOKALARTIKULATION. Von hier aus werden die unabhängigen akustischen Analysen bezüglich der Vokalartikulation oder der Stimme initiiert.

Bei konstanter Stellung der Artikulatoren sind die korrespondierenden Formantverläufe ebenfalls näherungsweise konstant.
Auf der rechten Seite ist das jeweils entsprechende mittlere LPC-Spektrum dargestellt. Diese Frequenzbereichs-Repräsentation des Sprachsignals entsteht durch Mittelung über sämtliche, im Abstand von 10 ms neuberechneten (Einzel-)LPC-Spektren; die gestrichelten Linien markieren die Minimal- bzw. Maximalwerte innerhalb dieser Menge. Die LPC-Spektren zeigen im Normalfall zwei markante Überhöhungen bei den beiden Formantfrequenzen F1 und F2.
Im Prinzip entsteht jedes Formantpaar der links dargestellten Formantverläufe aus den beiden detektierten Peaks des jeweils korrespondierenden LPC-Spektrums; der gesamte Verlauf ergibt sich aus der segmentweisen Verfolgung dieser Peaks vom Anfang bis zum Ende der Aufzeichnung.
Jedes Formant-Tracking-Verfahren, auch das in MODIAS implementierte, kann in Einzelfällen fehlerbehaftet sein oder völlig divergieren. Die Tracking-Startwerte (Wertepaar

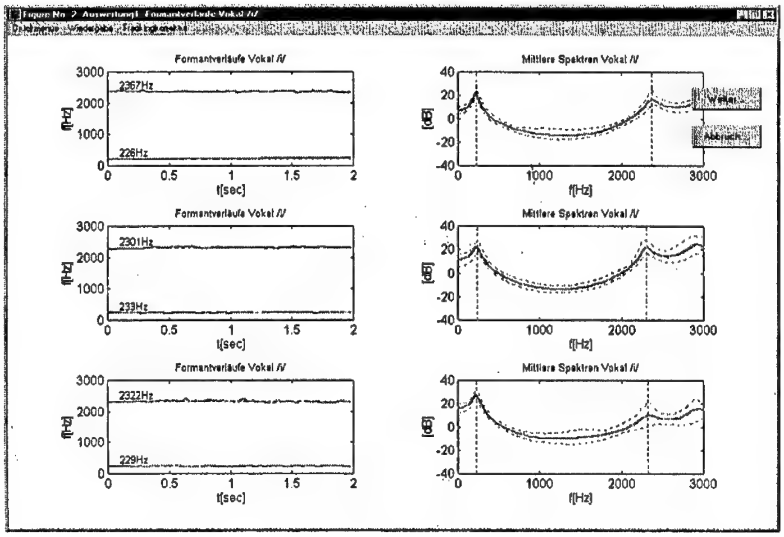

Abbildung 4.9.: Auswertung der Artikulationsanalyse zum Modul STIMME UND VOKALARTIKULATION, Blatt 1-4. Jeder der Vokale wird auf einem eigenen Auswertungsblatt repräsentiert (hier: */i/*) mit allen Realisierungen der Sitzung.

F1/F2 des ersten Signalsegments) sind hier besonders kritisch bezüglich der korrekten Funktionalität. Durch die LPC-Spektren wird dem Untersucher eine alternative Darstellung angeboten, um

- den (abgeleiteten) Formantverlauf bei vermuteten Störungen bewerten und
- gegebenenfalls den Formant-Tracking-Algorithmus steuernd beeinflußen zu können.

Das Menü TRACKINGKORREKTUR regelt die Interaktion des Untersuchers bei beobachteten Divergenzen des Formant-Tracking-Algorithmus. Solche Divergenzen liegen immer dann vor, wenn mindestens einer der beiden Formantverläufe

- unvermittelt nach oben oder unten ausbricht oder

- sich insgesamt auf einem nicht plausiblen Niveau bewegt

und der perzeptive Höreindruck eine gegensätzliche Erwartungshaltung induziert.
Die Mediane der beiden links dargestellten Formantverläufe werden grundsätzlich in dem rechts dargestellten, mittleren LPC-Spektrum markiert. Wurde, eventuell bedingt durch einen fehlerhaften Tracking-Startwert, ein falscher Peak des korrespondierenden LPC-Spektrums detektiert und dessen zeitlicher Verlauf anstelle des richtigen Peaks durch alle LPC-Spektren hindurch fehlerhaft verfolgt, kann hier bei der Auswahl MANUELLE VORGABE ein neues Wertepaar als Tracking-Startwert direkt im LPC-Spektrum vorgegeben werden.
Hierzu werden die beiden Startwerte (also die offensichtlich richtigen Peaks des LPC-Spektrums) jeweils mit der linken Mousetaste markiert. Nach kurzer Neuberechnung werden die beiden resultierenden Formantverläufe auf der linken Seite des Auswertungsblattes eingeblendet, die damit ebenfalls neuberechneten Mediane werden auf der rechten Seite im zugehörigen LPC-Spektrum aktualisiert.
Die Auswahl AUTOMATISCHES TRACKING gestattet die Rückkehr zu den ursprünglichen Verläufen und Werten, d.h. zum vollautomatischen Tracking ohne nutzerspezifische Startwerte.
Das Menü DRUCKEN stellt diverse Varianten der Druckausgabe zur Verfügung wie Farbdruck, Schwarz-Weiß-Druck oder Kopie in die Zwischenablage für die weitere Verwendung in klinischen Berichten usw.
Das Menü WIEDERGABE erlaubt die akustische Wiedergabe der jeweils zugehörigen Sprachsignale bei gleichzeitiger Betrachtung der ermittelten Parameterverläufe. Diese Art der Verknüpfung von akustischer Information und perzeptivem Höreindruck soll den Untersucher in seiner Analyse möglichst optimal unterstützen.
Das Beispiel gem. Abbildung 4.9 zeigt die 3 Realisierungen des Vokals */i/* eines sprechgesunden Probanden. Sowohl die Formantverläufe als auch die korrespondierenden LPC-Verläufe zeigen eine geringe Formant-Variabilität und weisen damit auf eine sehr konstante Stellung der Artikulatoren hin. Der Vergleich der beiden Formantverläufe (Mediane) einerseits, und der entsprechenden LPC-Spektren (Peaks) andererseits läßt auf Konvergenz und damit auf eine optimale Genauigkeit des Formant-Trackings schließen.
Das anschließende Auswertungsblatt gem. Abbildung 4.10 faßt die Ergebnisse der Artikulationsanalyse zusammen und stellt diese dem Untersucher in übersichtlicher Form dar.

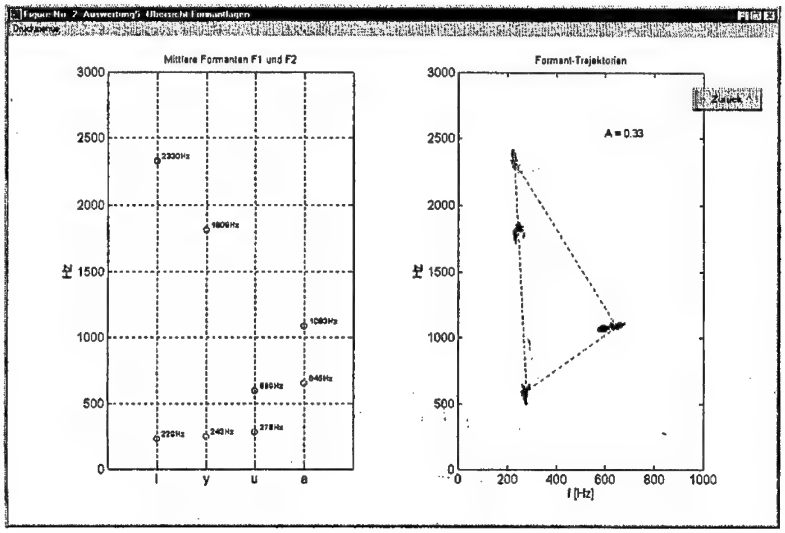

Abbildung 4.10.: Auswertung der Artikulationsanalyse zum Modul STIMME UND VOKALARTIKULATION, Blatt 5. Die Zusammenfassung sämtlicher Ergebnisse der Artikulationsanalyse bietet dem Untersucher einen Gesamtüberblick über die mittleren Formantlagen F1/F2 aller 4 Vokale (links), sowie über die Öffnung des Artikulationsraumes anhand einer Formantkarte (rechts).

Auf der linken Bildhälfte sind die mittleren Formanten F1 und F2 der 4 Vokale (hier gemittelt über alle Realisierungen des betreffenden Vokals) entlang der Frequenzachse dargestellt. Auf der rechten Seite werden die mittleren Formanten in eine 2-dimensionale F1/F2-Ebene bzw. Formantkarte eingetragen. Die Vokale */i/*, */u/* und */a/* bilden das sogenannte *Formantdreieck*, dessen Fläche ein sensitives Maß für den Artikulationsraum bzw. dessen Öffnung darstellt und daher in der grafischen Ausgabe berücksichtigt wurde (Dimension kHz^2).

Zusätzlich werden alle (bis zu 12) zeitlichen Formantverläufe in diese Ebene projiziert. Diese gewählte Art der Darstellung erlaubt auf einen Blick die Beurteilung der zeitlichen Stabilität der artikulatorischen Konfiguration und liefert unmittelbar eine Aussage über

die bei Patienten mit neurogenen Sprechstörungen häufig beobachteten

- zeitlichen Formantverschiebungen in Richtung anderer Vokale, besonders aber über
- Zentralisierungstendenzen bei Erschlaffung der am Sprechvorgang beteiligten Muskulatur (Vokale klingen mit zunehmender Realisierungsdauer gleich).

4.2.3.2. Auswertung der Stimmqualität und Stimmstabilität

Die Stimmanalyse basiert auf dem gleichen Vorrat an Vokalaufzeichnungen wie die bereits beschriebene Artikulationsanalyse, jedoch werden hier nur die beiden Vokale */i/* und */a/* herangezogen. Sie repräsentieren die beiden Extremstellungen des Kehlkopfs, der durch die vorgelagerte und angehoben Zunge beim */i/* gekippt und angehoben wird, beim */a/* dagegen abgesenkt und entspannt ist. Die beiden Vokale sind daher vom diagnostischen Standpunkt aus relevant.
Die ersten beiden Auswerteblätter gem. Abbildung 4.11 widmen sich dem Verlauf der Stimmbandgrundfrequenz (Pitch) durch Analyse jeder Einzelschwingung der zugehörigen Sprachsignale. Den beiden Vokalen ist jeweils ein eigenes Auswertungsblatt zugeordnet mit allen Realisierungen dieser Sitzung. Auf der linken Seite sind stets die zeitlichen Pitchverläufe sowie deren Mediane dargestellt.
Bei konstantem Stimmbandtonus und konstanter Stimmlippenadduktion sind die korrespondierenden Pitchverläufe ebenfalls näherungsweise konstant.
Auf der rechten Seite ist ein Ausschnitt aus dem jeweils entsprechenden Oszillogramm, also dem zeitlichen Schalldruckverlauf des Sprachsignals, dargestellt. Im Prinzip entsteht jeder einzelne Wert des links dargestellten Pitchverlaufs aus dem Abstand zweier benachbarter Schwingungsmaxima (Periodendauer) des Oszillogramms, der gesamte Verlauf aus der periodenweisen Verfolgung dieser Maxima vom Anfang bis zum Ende der Aufzeichnung. Jedes Pitch-Tracking-Verfahren, auch das in MODIAS implementierte, kann in Einzelfällen fehlerbehaftet sein oder völlig divergieren. Der Startwert (erste gemessene Periodendauer) ist besonders kritisch bezüglich der korrekten Funktionalität des Tracking-Algorithmus.
Durch Darstellung der Oszillogramme wird dem Untersucher eine alternative Darstellung angeboten, um

- den Pitchverlauf bei vermuteten Störungen bewerten und

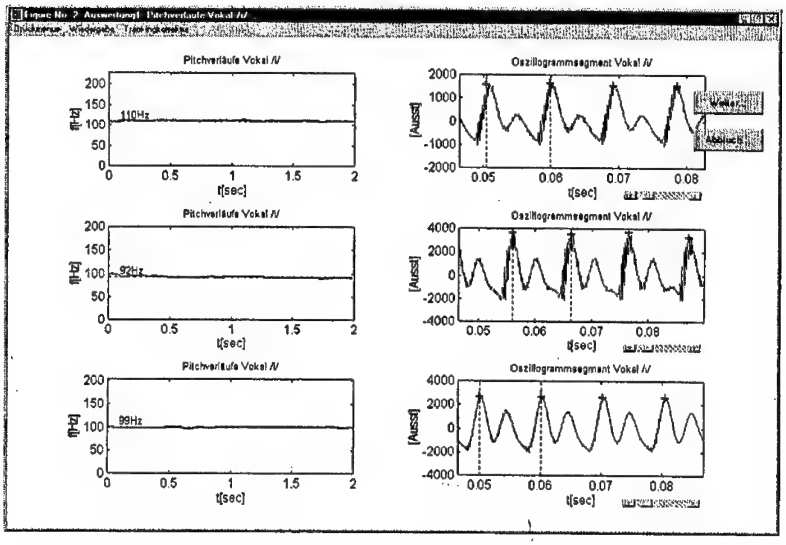

Abbildung 4.11.: Auswertung der Stimmanalyse des Moduls STIMME UND VOKALARTIKULATION, Pitchverläufe, Blatt 1-2 (hier: Vokal /*i*/).

- gegebenenfalls den Pitch-Tracking-Algorithmus steuernd beeinflussen zu können.

Das Menü TRACKINGKORREKTUR regelt die Interaktion des Untersuchers bei beobachteten Divergenzen des Pitch-Tracking-Algorithmus. Solche Divergenzen liegen immer dann vor, wenn der Pitchverlauf

- unvermittelt nach oben oder unten ausbricht oder
- sich insgesamt auf einem nicht plausiblen Niveau bewegt

und der perzeptive Höreindruck eine gegensätzliche Erwartungshaltung induziert.
Die detektierten Schwingungsmaxima werden grundsätzlich in dem rechts dargestellten Oszillogramm markiert (Pitch-Marken). Wurde, eventuell bedingt durch einen fehlerhaften Tracking-Startwert, zum Beispiel nur jedes zweite Maximum detektiert, kann hier durch die Auswahl MANUELLE VORGABE ein neues Wertepaar als Tracking-Startwert

direkt im Oszillogramm vorgegeben werden.
Hierzu werden die beiden Startwerte (also zwei aufeinanderfolgende Pitch-Marken im Abstand einer Periodendauer) jeweils mit der linken Mousetaste markiert. Nach kurzer Neuberechnung wird der resultierende Pitchverlauf auf der linken Seite des Auswertungsblattes eingeblendet, alle neuberechneten Pitch-Marken werden auf der rechten Seite im zugehörigen Oszillogramm aktualisiert.
Die Auswahl AUTOMATISCHES TRACKING gestattet die Rückkehr zu den ursprünglichen Verläufen und Werten, d.h. zum vollautomatischen Tracking ohne nutzerspezifische Startwerte.
Die Menüs DRUCKEN und WIEDERGABE sind identisch zu denen des vorhergehenden Abschnittes (Auswertung der Vokalartikulation).
Die zwei folgenden Auswertungsblätter gem. Abbildung 4.12 widmen sich dem Verlauf des komplexen Parameters CPPS (Cepstral Peak Prominence Smoothed), der besonders mit der perzeptiven Stimmqualität *Behauchtheit*, aber auch mit der Stimmqualität *Rauhigkeit* korreliert. Beide Stimmerkmale können in ihrem zeitlichen Verlauf über die Dauer der Vokalartikulation hinweg allein auf der Basis akustischer Analysen bewertet werden.
Den beiden Vokalen ist jeweils ein eigenes Auswertungsblatt zugeordnet mit allen Realisierungen dieser Sitzung.
Links sind die zeitlichen Verläufe der CPPS einschließlich ihrer Mediane dargestellt.
Auf der rechten Seite ist das jeweils entsprechende hochauflösende und gemittelte Schmalband-Leistungsdichtespektrum dargestellt.
Im Prinzip entsteht jeder der im Abstand von 10 ms neuberechneten CPPS-Werte des CPPS-Verlaufs aus einem entsprechenden (Einzel-)Leistungsdichtespektrum. Die Spektren zeigen bei hoher Stimmqualität eine erkennbar periodische Feinstruktur (Spektrallinien), deren Abstand und Verteilung unmittelbar vom Pitch abhängt.
Eine wahrnehmbare Rauhigkeit korrespondiert physiologisch mit einem Phasenjitter der Stimmbandanregung, eine Behauchtheit mit einer dieser Stimmbandschwingung unterliegenden Rauschkomponente. Beide Effekte führen zu einer Störung der periodischen Feinstruktur des Leistungsdichtespektrums und damit zu einer Absenkung des CPPS-Parameters.
Das Beispiel gem. Abbildung 4.12 zeigt die resultierenden Ergebnisse für eine synthetische Vocoder-Stimme

- ohne Stimmbandjitter (obere Grafiken);

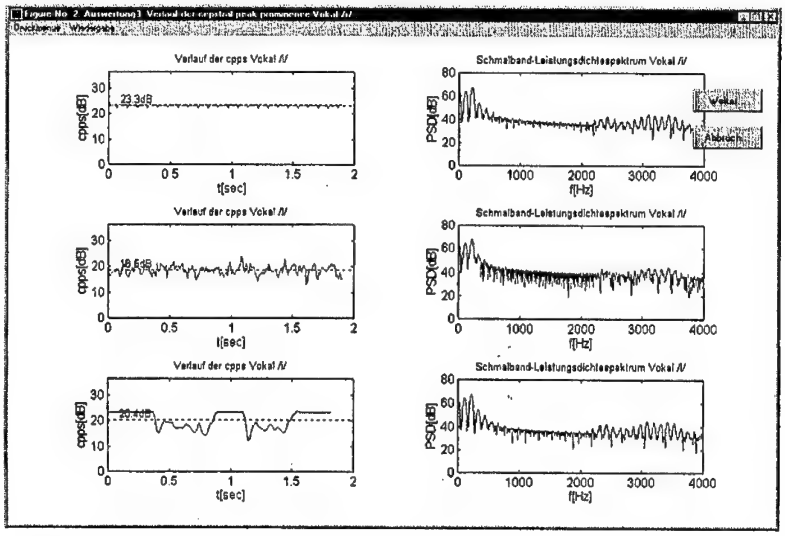

Abbildung 4.12.: Auswertung der Stimmanalyse des Moduls STIMME UND VOKALARTIKULATION, CPPS-Verläufe, Blatt 1-2.

- mit zeitkontinuierlichem Random-Jitter von 1.0 Prozent (d.h. relative Änderung des Pitchs von einer Schwingung zur nächsten beträgt maximal 1 Prozent). Die Rauhigkeit ist perzeptiv kaum wahrnehmbar, der CPPS-Parameter reagiert jedoch hochsensitiv und ist um ca. 5 dB vermindert (mittlere Grafiken);
- mit zeitvariablem Random-Jitter. Zunächst ohne, dann aber mit abrupt intermittierendem Random-Jitter von 5.0 Prozent in zwei Phasen. Die Rauhigkeit ist perzeptiv deutlich wahrnehmbar, der CPPS-Parameter bildet die Übergänge zu den Phasen mit simulierten Stimmstörungen durch Verlaufseinbrüche um bis zu 10 dB ab.

Das jetzt folgende Auswertungsblatt gem. Abbildung 4.13 betrachtet zusammenfassend den Teilaspekt der Stimmqualität. Es stellt dem Untersucher die zur Bewertung relevanten Informationen in übersichtlicher Form dar.
Das obere Drittel des Auswertungsblattes zeigt im linken Teil die mittlere Tonhöhe (Pitch) der jeweils 3 Realisierungen pro Vokal, im rechten Teil die Gesamtmittelwerte

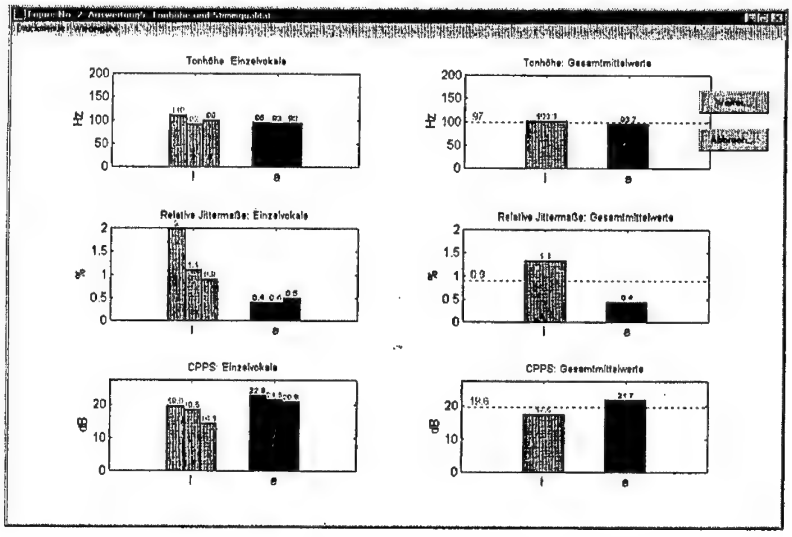

Abbildung 4.13.: Auswertung der Stimmanalyse des Moduls STIMME UND VOKALARTIKULATION mit den zusammenfassenden Ergebnissen zu den Aspekten Tonhöhe und Stimmqualität.

für die beiden betrachteten Vokale */i/* und */a/* sowie deren Gesamtmittelwert (Grand Average). Das mittlere Drittel des Auswertungsblattes stellt analog im linken Teil die mittleren relativen Jittermaße (MODIAS-spezifische Jitterdefinition, nicht identisch zu der Definition bei der synthetischen Signalerzeugung des beschriebenen Vocoders) aller realisierten Vokale dar und im rechten Teil wieder die Gesamtmittelwerte und den Grand Average.
Im unteren Drittel des Auswertungsblattes sind links die mittleren CPPS-Werte und rechts die entsprechenden Gesamtmittelwerte dargestellt.

Das letzte Auswertungsblatt gem. Abbildung 4.14 betrachtet zusammenfassend den Teilaspekt der Stimmstabilität. Während bei der Interpretation der Stimmqualität die mittleren (Absolut-)Lagen von Tonhöhe, Jitter und CPPS relevant waren, interessieren hier nun die über den Zeitraum der Vokalisation aufgetretenen relativen Schwankungen

von Tonhöhe, Lautheit und CPPS. Zur Ableitung des Schwankungsparameters für die Lautheit werden die bei der Segmentierung verwendeten Lautheitsverläufe herangezogen.

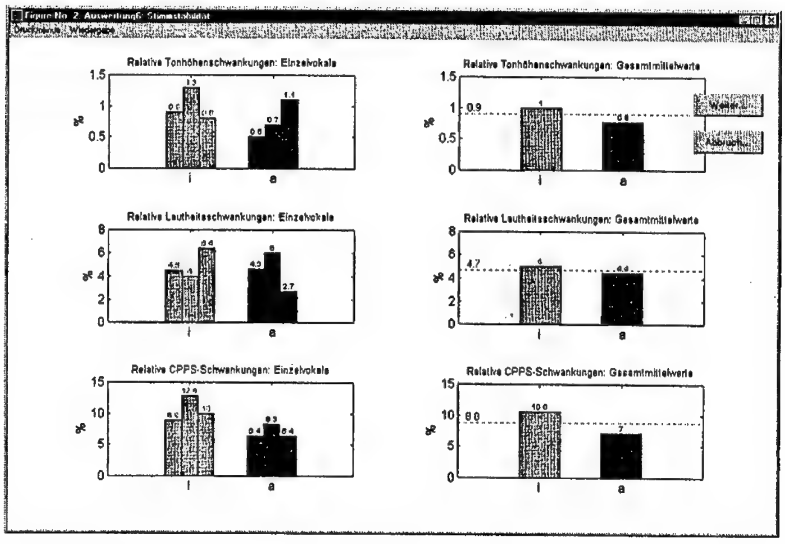

Abbildung 4.14.: Auswertung der Stimmanalyse des Moduls STIMME UND VOKALARTIKULATION mit den zusammenfassenden Ergebnissen zu dem Aspekt Stimmstabilität.

Die Darstellungslogik wurde analog zu dem beschriebenen Auswertungsblatt gewählt, das vorhergeht.

4.2.3.3. Automatischer Datenexport

Mit Aufruf der Auswertung und damit synchron zur Darstellung des ersten Auswertungsblattes gem. Abbildung 4.9 bzw. 4.11 wird automatisch je 1 Datenfile mit allen relevanten Ergebnissen im entsprechenden Arbeitsverzeichnis angelegt:

- Artikulationsdatenfile: Datenfile mit der Endung *...art.asc*; es enthält die mittleren

Formanten F1 bzw. F2 für jeden Vokal.

- Stimmdatenfile: Datenfile mit der Endung *...sti.asc*; es enthält die Ergebnisse statistischer Analysen der Zeitverläufe von Pitch und Lautheit für jeden Vokal.

Beide Datenfiles sind ASCII-codiert und damit universell zur weiteren Nutzung oder Weiterverarbeitung mit alternativen Softwareprodukten einsetzbar.

Format des Artikulationsdatenfiles: Das Artikulationsdatenfile enthält ein Zahlenfeld in Matrixdarstellung, wobei jedem Vokal eine Matrixspalte zugeordnet ist. Die Belegung der Spalten erfolgt gem. Tabelle 4.1.
Jede Zeile enthält folgende Informationen:

- Zeile 1: Mittlerer Formant F1,
- Zeile 2: Mittlerer Formant F2.

Matrixspalte	Vokal	Realisierung
1	/i/	1
2	/y/	1
3	/u/	1
4	/a/	1
5	/i/	2
6	/y/	2
7	/u/	2
8	/a/	2
9	/i/	3
10	/y/	3
11	/u/	3
12	/a/	3

Tabelle 4.1.: Identifizierung des Vokals anhand der Spaltennummer im Artikulationsdatenfile.

Das Artikulationsdatenfile bzw. die enthaltene Datenmatrix verfügt also über eine feste Anzahl von 2 Zeilen und 12 Spalten.

Nicht realisierte Vokale werden durch die Werte -1 in den zugeordneten Spalten der Matrix gekennzeichnet.

Format des Stimmdatenfiles: Das Stimmdatenfile enthält ebenfalls ein Zahlenfeld in Matrixdarstellung, wobei jeder Matrixzeile einem Vokal fest zugeordnet ist gem. Tabelle 4.2.

Matrixzeile	Vokal	Realisierung
1	/i/	1
2	/i/	2
3	/i/	3
4	/a/	1
5	/a/	2
6	/a/	3

Tabelle 4.2.: Identifizierung des Vokals anhand der Zeilennummer im Stimmdatenfile.

Jede Zeile dieser Matrix enthält in 9 Feldern (Spalten) die folgenden Informationen (erläutert in 4.3.2.1 auf Seite 93, 4.3.2.2 auf Seite 96 und 4.3.2.3 auf Seite 104):

- Feld 01: Jitter in Prozent,
- Feld 02: Pitch-Variabilität in *Prozent über Trend*,
- Feld 03: Pitch-Trend in *Prozent pro Sekunde* (positives Vorzeichen bei ansteigendem Pitch),
- Feld 04: Mittlerer Pitch in *Hz*,
- Feld 05: Lautheits-Variabilität in *Prozent über Trend* (positives Vorzeichen bei Zunahme der Lautheit),
- Feld 06: Lautheits-Trend in *Prozent pro Sekunde*,
- Feld 07: CPPS-Variabilität in *Prozent über Trend* (positives Vorzeichen bei Zunahme der CPPS),

- Feld 08: CPPS-Trend in *Prozent pro Sekunde*,
- Feld 09: Mittlere CPPS in *dB*.

Die akustischen Parameter erlauben eine detaillierte Bewertung von 3 verschiedenen Instabilitätsarten eines Zeitverlaufs. Siehe hierzu auch Abbildung 4.29 auf Seite 92. Das Stimmdatenfile bzw. die darin enthaltene Datenmatrix verfügt über eine feste Anzahl von 9 Zeilen und 12 Spalten.

4.3. Technische Realisierung: Entwickelte Verfahren und Algorithmen

4.3.1. Verfahren zur Prüfung des Artikulationsraumes

4.3.1.1. Vokaltraktmodellierung und Formantanalyse

Zur Prüfung des Resonanzraumes und der Stellung der Artikulatoren ist eine geeignete Modellierung des Vokaltraktes mit seinen Sprechwerkzeugen erforderlich. Ein solches Ersatzmodell kann auf Grund der Zeitvarianz des Sprachsignales natürlich nur innerhalb kurzer Signalabschnitte Gültigkeit besitzen, so daß mit hinreichender Genauigkeit von quasistationären Verhältnissen ausgegangen werden kann. Bedingt durch physiologische und physikalische Gegebenheiten setzt dies in praxi die Berechnung von ca. 100 solcher Modelle pro Sekunde voraus. Die nachfolgend dargestellte Ableitung zeigt den im Modul STIMME UND VOKALARTIKULATION beschrittenen technischen Weg von der Aufzeichnung des Sprachsignales über die Vorverarbeitung und Vokaltraktmodellierung bis hin zur Ermittlung der zeitlichen Verläufe der beiden Formanten F1 und F2.

Decimation und Preemphasis: Das aufgezeichnete und in digitaler Repräsentation vorliegende Sprachsignal (Samplingrate 22050 Hz, Auflösung 16 Bit) wird zunächst in zwei nacheinander ausgeführten Schritten vorverarbeitet, das ist die *Decimation* und die *Preemphasis*.
Die zu erwartenden Formantfrequenzen sind grundsätzlich bei dem Vokal */i/* am höchsten, sie liegen jedoch sicher unterhalb einer Frequenz von ca. 3000 Hz. Damit kann die Samplingrate um den Faktor 2 reduziert werden, ohne informationstragende Signalanteile zu eliminieren. Die hier realisierte Reduktion der Samplingrate (Decimation) richtet sich nach dem Decimation-Algorithmus gem. IEEE ([IEEE79]) und erfolgt in 2 Stufen:

- Tiefpaßfilterung des Sprachsignales,
- Extraktion nur noch jedes 2. Samples.

Das Tiefpaßfilter ist nach [IEEE79] als Chebyshev-Filter 8.Ordnung ausgeführt mit 0.05 dB Ripple im Durchlaßbereich. Der zugehörige Amplituden- und Phasengang des

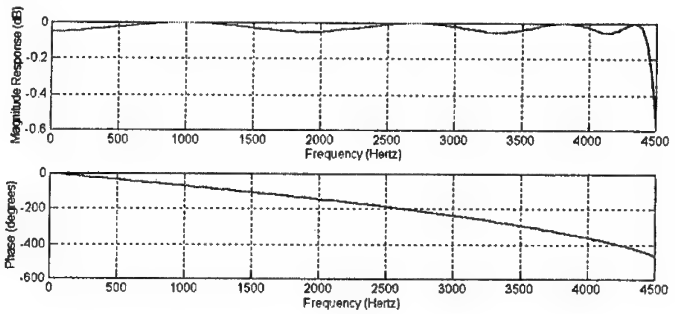

Abbildung 4.15.: Amplituden- und Phasengang des verwendeten Decimation-Filters (Chebyshev-Filter 8. Ordnung mit 0.05 dB Ripple im Durchlaßbereich nach IEEE).

Decimation-Filters ist in Abbildung 4.15 dargestellt.
Nach Abschluß der Decimation ist die Samplingrate um den Faktor 2 reduziert, sie beträgt daher für alle weiteren Betrachtungen nur noch 11025 Hz.
Die folgende Hochpaßfilterung (Preemphasis) kompensiert weitgehend den Tiefpaßeffekt der physiologischen Abstrahlung des Sprachschalles über Mund und Lippen, der höhere Formanten bedämpfen und somit zu Störungen in den zeitlichen Formantverläufen führen kann. Hierfür wurde ein einfaches Moving-Average bzw. FIR-Filter gewählt, da diese Filtertypen die Eigenschaft der exakten Linearphasigkeit besitzen und damit unerwünschte Phasenverzerrungen vermeiden.
Die Systemfunktion des Preemphasisfilters wurde wie folgt gewählt (First-Difference-Filter):

$$H_{Pre}(z) = b_1 z^{-1} + b_2 z^{-2} = 1z^{-1} - 1z^{-2} \tag{4.1}$$

Der zugehörige Amplituden- und Phasengang ist in Abbildung 4.16 dargestellt.

Segmentbildung: Aus dem vorverarbeiteten Sprachsignal werden nun Ersatzmodelle für den Vokaltrakt berechnet, die für kurze Zeiträume (hier: 10ms) Gültigkeit besitzen. Die Parameter dieser Modelle - hier die Filterkoeffizienten des digitalen Vokaltraktfilters - geben Aufschluß über die tatsächlichen Eigenschaften des realen Vokaltraktes des

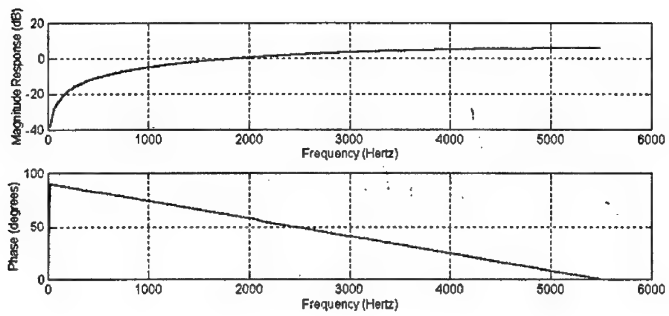

Abbildung 4.16.: Amplituden- und Phasengang des verwendeten Preemphasis-Filters (FIR-Filter 2. Ordnung)

Patienten, die eine sehr exakte Aussage über die für Vokale charakteristischen Resonanzfrequenzen (Formanten) und damit letztlich über die korrekte Stellung der Artikulatoren zulassen.
Die Modellierung des Vokaltraktes erfolgt über das Prinzip der linearen Prädiktion (Linear Predictive Coding LPC), welche eine Kurzzeitanalyse bzw. die Bildung zeitlich begrenzter Segmente bedingt. Hierzu wird bei der konkreten MODIAS-Realisierung das gesamte vorverarbeitete Sprachsignal in überlappende Segmente zu je 256 Samples aufgeteilt. Die Startpunkte aufeinanderfolgender, überlappender Segmente liegen in einem zeitlichen Abstand von 10ms, damit ist auch die zeitliche Auflösung der späteren Formantverläufe festgelegt auf 100 Werte pro Sekunde. Die gewählten Werte für Segmentlänge und Überlappung finden sich in diversen Publikationen (z.B. in [Rab78]) und werden hier ohne weitere Prüfung übernommen.

Vokaltraktmodellierung durch LPC-Analyse: Die folgenden Erläuterungen zeigen das realisierte Verfahren der linearen Prädiktion zur Ermittlung der Koeffizienten des Vokaltraktfilters auf. Die Ableitung ist im wesentlichen [Mark72] und [Mark73] entnommen.
Ausgangspunkt für alle Folgebetrachtungen ist das bekannte Quelle-Filter-Modell der Digitalen Sprachverarbeitung gem. Abbildung 4.17.
Die verschiedenen Methoden der linearen Prädiktion setzen als Annahme voraus, daß jeder Sample des Sprachsignales $s(n)$ durch eine Linearkombination vorausgehender Samp-

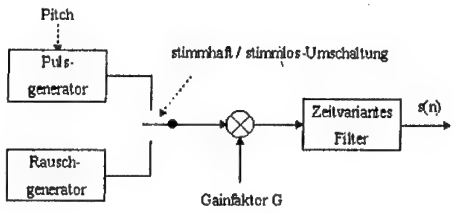

Abbildung 4.17.: Quelle-Filter-Modell der Digitalen Sprachsignalverarbeitung. Die Anregung durch die Stimmbänder wird hier durch einen Pulsgenerator (bei stimmhafter Anregung) bzw. durch einen Rauschgenerator (bei stimmloser Anregung) modelliert. Die Anregungsintensität wird über den variablen Gainfaktor G berücksichtigt. Die Artikulation durch den Vokaltrakt wird durch ein zeitvariantes Digitales Filter nachgebildet.

les näherungsweise in seiner Amplitude vorhergesagt werden kann. Ein linearer Prädiktor vom Grad K versucht also, aus den vorangegangenen $k = 1 \ldots K$ Abtastwerten des Sprachsignales den gerade aktuellen Abtastwert $s(n)$ durch eine möglichst erwartungstreue Schätzung $\tilde{s}(n)$ vorherzusagen.

$$\tilde{s}(n) = a_1 s(n-1) + a_2 s(n-2) + \ldots + a_K s(n-k) = \sum_{k=1}^{K} a_k s(n-k) \qquad (4.2)$$

Die Schätzung jedes einzelnen Samples ist fehlerbehaftet mit dem Prädiktionsfehler bzw. Prädiktions-Restsignal $e(n)$:

$$s(n) = \tilde{s}(n) + e(n) = \sum_{k=1}^{K} a_k s(n-k) + e(n) \qquad (4.3)$$

Nach einer z-Transformation beider Seiten der Gleichung 4.3 stellt sich der Zusammenhang wie folgt dar:

$$S(z) = \sum_{k=1}^{K} a_k S(z) z^{-k} + E(z) \qquad (4.4)$$

Beziehungsweise gilt nach Umformung:

$$S(z) = E(z) \frac{1}{1 - \sum_{k=1}^{K} a_k z^{-k}} \quad \text{bzw.} \quad S(z) = E(z) H_{vok}(z) \quad \text{mit} \qquad (4.5)$$

$$H_{vok}(z) = \frac{1}{1 - \sum_{k=1}^{K} a_k z^{-k}} \tag{4.6}$$

Durch Einführung der neuen Systemfunktion $H_{vok}(z)$ läßt sich der Zusammenhang gem. Gleichung 4.6 interpretieren als ein LZI-System mit Systemfunktion $H_{vok}(z)$, das durch den Prädiktionsfehler $e(n)$ angeregt wird und am Ausgang das tatsächliche Sprachsignal $s(n)$ produziert.

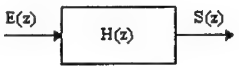

Abbildung 4.18.: Interpretation der Gleichung 4.6 als ein LZI-System, das durch das Prädiktionsfehlersignal $e(n)$ angeregt wird und am Ausgang das ursprüngliche Sprachsignal $s(n)$ generiert.

Je besser dabei die Prädiktion gelingt, desto stärker wird das Prädiktionsfehlersignal dekorreliert, d.h. die statistische Bindung zwischen den Abtastwerten reduziert. Das Fehlersignal nähert sich bei optimaler Prädiktion idealerweise weißem Rauschen, daher wird diese Filter auch als *Spectral-Whitening-Filter* bezeichnet.

Der Zusammenhang mit dem Quelle-Filter-Modell wird jetzt offensichtlich: Die Systemfunktion $H_{vok}(z)$ schätzt die Filtereigenschaften des realen Vokaltraktes innerhalb kurzer Zeitsegmente unter der Annahme von Quasistationarität, da das reale Stimmbandsignal im Falle stimmhafter wie auch im Falle stimmloser Anregung ein rauschähnliches Spektrum und damit (rausch-)ähnliche Eigenschaften wie das Prädiktionsfehlersignal aufweist.
Im (realen) Falle endlicher Prädiktionsgüte und geeignet gewählter Prädiktorlänge (hier: K=16) stellt das resultierende Prädiktionsfehlersignal der linearen Prädiktion eine hinreichend gute Schätzung für die Stimmbandanregung dar.
Das zu lösende Problem liegt nun darin, die Prädiktorkoeffizienten derart zu wählen, daß der Prädiktionsfehler bezogen auf eine definierte Anzahl von Samples minimal wird. Wählt man hierfür als Kriterium den mittleren quadratischen Fehler (Least-Mean-Square-Kriterium) bzw. die Energie des betrachteten Segmentes, so führt das Minimierungsproblem zu mathematisch lösbaren Gleichungssystemen.

$$E = \sum_{n} e^2(n)! = Min \tag{4.7}$$

Die Bestimmung der gesuchten Prädiktorkoeffizienten a_k erfolgt in dem betrachteten MODIAS-Modul nach der Autokorrelationsmethode (eine detailliertere Beschreibung hierzu findet sich in [Mark72]).
Dieser Lösungsansatz setzt einschränkend voraus, daß das Sprachsignal $s(n)$ außerhalb des Intervalles $n \in [0...N-1]$ identisch Null ist bzw. mit einer entsprechenden Rechteck-Fensterfunktion multipliziert wird. Damit liefert der Prädiktionsfehler e(n) nur innerhalb des Intervalles $n \in [0...N-1+K]$ Werte ungleich Null, und das Minimierungsproblem aus Gleichung 4.7 vereinfacht sich zu:

$$E = \sum_{n=0}^{N-1+K} e^2(n)! = Min \tag{4.8}$$

Der Vergleich mit Gleichung 4.3 zeigt, daß der Prädiktionsfehler $e(n)$ bei der Autokorrelationsmethode hohe Werte am Anfang des betrachteten Segmentes $n \in [0...K-1]$ liefern muß, da der Prädiktor hier eine Schätzung von Samples durch Linearkombination aus einer Reihe von Nullen erzwingt. Das gleiche gilt für das Ende des betrachteten Segmentes $n \in [N...N-1+K]$; hier werden umgekehrt Nullen aus einer Reihe von Samples (ungleich Null) gewonnen.
Aus diesem Grund ist für die Autokorrelationsmethode eine geeignete Fensterfunktion sinnvoll, die zum Rand des Segmentes auf Null abfällt und damit diesen Effekt auszugleichen versucht. Im beschriebenen Anwendungsfall wurde hierzu ein Hamming-Fenster gewählt. Die Elemente des Hamming-Fensters können wie folgt mathematisch beschrieben werden:

$$w(n) = 0.54 - 0.46\cos(2\pi\frac{n}{N-1}) \quad \text{für alle} \quad n \in [0 \ldots N-1] \tag{4.9}$$

Die aus Gleichung 4.9 resultierende Form des Hamming-Fensters ist in Abbildung 4.19 dargestellt.
Das aus Gleichung 4.8 resultierende Gleichungssystem ist z.B. in [Mak72] ausführlich beschrieben. Die Lösung dieses Gleichungssystems führt zu dem gesuchten Koeffizientensatz $a_1, a_2 \ldots a_k$.
MODIAS zieht hier die rechenzeitsparende Levinson-Durbin-Rekursion heran, die u.a. in [Rab78] angegeben ist und hier nicht weiter erläutert wird.

Auswahl resonanzfähiger Pole und Berechnung der Formantfrequenzen: Das so erhaltene Digitale Filter mit der Systemfunktion $H_{vok}(z)$ enthält in der komplexen z-Ebene

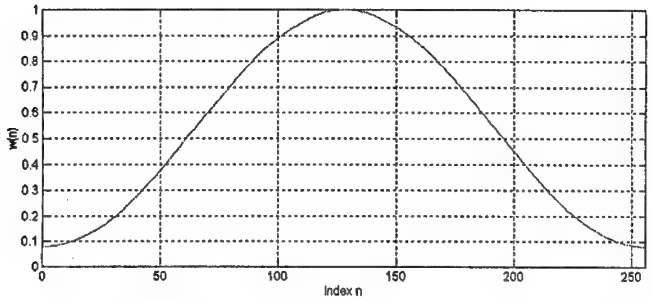

Abbildung 4.19.: Form des gewählten Hamming-Fensters. In MODIAS wird jedes Signalsegment $s(n)$ der Länge 256 vor der LPC-Analyse mit einem Hamming-Fenster multipliziert, um so den entstehenden Prädiktionsfehler zu den Segmentgrenzen hin nicht übermäßig ansteigen zu lassen.

nur reelle oder konjugiert komplexe Polpaare, jedoch keine Nullstellen (AR-Filter bzw. Allpol-Modell). Dieses Filtermodell eignet sich daher hervorragend für Vokale, da es die für diese Lautklasse typischen Überhöhungen in der zur Systemfunktion korrespondierenden Übertragungsfunktion $|H_{vok}(j\omega)|$ prinzipiell zu erzeugen vermag. Dabei tragen die konjugiert komplexen Polpaare die Information über die Lage (Frequenz) und über die Ausprägung (Bandbreite) dieser spektralen Resonanzen.
Die Polstellen werden im beschriebenen MODIAS-Modul analytisch durch Nullstellenberechnung des Nennerpolynoms der Systemfunktion $H_{vok}(z)$ gewonnen (*Root-Solving-Verfahren*). Damit gilt:

$$A(z) = \frac{1}{H_{vok}(z)} = 1 - \sum_{k=1}^{K} a_k z^{-k} = \prod_{k=1}^{K} (1 - z_k z^{-1}) \quad != 0 \tag{4.10}$$

Der Zusammenhang zwischen einem konjugiert komplexen Polpaar (z_i, z_i^*) der Systemfunktion $H_{vok}(z)$ und zugehöriger Resonanzfrequenz und -bandbreite im Fourierspektrum kann durch Übergang von der z-Transformation zur Laplace-Transformation hergestellt werden, also durch Abbildung des komplexen Einheitskreises auf die komplexe Frequenzebene. Die Umrechnung auf Frequenz F_i und Bandbreite B_i wird in der einschlägigen

Literatur (z.B. in [Rab78]) wie folgt angegeben:

$$F_i = \frac{F_s}{2\pi}\mathrm{arc}(z_i) \qquad B_i = -\frac{F_s}{\pi}\ln(|z_i|) \tag{4.11}$$

Problematisch ist bei jeder Formantanalyse die Zuordnung der so gewonnenen Resonanzfrequenzen zu den tatsächlichen Formanten, denn nicht jedes konjugiert komplexe Polpaar erzeugt auch eine *sichtbare Überhöhung* im Spektrum. Während die in der Literatur angegebenen Verfahren an dieser Stelle bei der Berechnung der Formantfrequenzen Näherungen benutzen und zusätzlich Plausibilitätskriterien unter Abstützung auf Erfahrungswerte bei Normsprechern heranziehen, kann bei sprechgestörten Patienten nicht auf derartige Normwerte zurückgegriffen werden.

Aus diesem Grund wurde zur Realisierung des MODIAS-Moduls ein Verfahren entwickelt, um sowohl

- Formanten von *Spectral-Shaping-Poles* algorithmisch unterscheiden und
- Formanten in ihrer Frequenzlage höher auflösen

zu können. Für die nun folgende Betrachtung wird die Systemfunktion $H_{vok}(z)$ des Vokaltraktfilters in Übertragungssysteme 2. Ordnung zerlegt, also in grundsätzlich resonanzfähige Subsysteme mit je einem konjugiert komplexen Polpaar:

$$H_{vok}(z) = \frac{1}{1-\sum_{k=1}^{K} a_k z^{-k}} = \frac{1}{\prod_{k=1}^{K}(1-z_k z^{-1})} = H_1(z)\cdot H_2(z)\cdot H_3(z)\ldots \tag{4.12}$$

Die Systemfunktion $H_{vok}(z)$ kann also bei gerader Anzahl von Polpaaren bzw. gerader Prädiktorlänge K als Kettenschaltung von $0.5K$ Teilsystemfunktionen 2. Ordnung $H_i(z)$ dargestellt werden.
Die Teilsystemfunktion $H_i(z)$ mit einem komplexen Polpaar (z_i, z_i^*) lautet:

$$H_i(z) = \frac{1}{(1-z_i z^{-1})(1-z_i^* z^{-1})} \tag{4.13}$$

Die Lage des Polpaares stellt sich dann in der komplexen z-Ebene gem. Abbildung 4.21 (linke Bildhälfte) dar.
Dabei korrespondieren die kartesische und die polare Darstellung der komplexen Größe z_i gem. folgendem Zusammenhang:

$$z_i = \mathrm{Re}(z_i) + j\mathrm{Im}(z_i) = |z_i|e^{j\cdot\mathrm{arc}(z_i)} \tag{4.14}$$

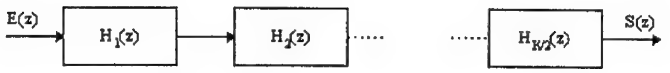

Abbildung 4.20.: Zerlegung der ursprünglichen, den menschlichen Vokaltrakt modellierenden Systemfunktion $H_{vok}(z)$ in eine Kette von (prinzipiell resonanzfähigen) Teilsystemfunktionen 2. Ordnung

Der Übergang von der Teilsystemfunktion $H_i(z)$ mit komplexer Polstelle z_i zur entsprechenden Laplace-Transformierten $H_i(s)$ mit komplexer Polstelle s_i läßt sich durch Einführung der in der Regelungstechnik bekannten Größen wie *Dämpfungsgrad* d , *Kennkreisfrequenz* ω_o und *Eigenkreisfrequenz* ω_e (Einführung in Gleichung 4.16) anschaulicher darstellen (in Anlehnung an [Boe88]).

Damit lautet die Laplace-Darstellung:

$$H_i(s) = \frac{1}{(1 + \frac{2d}{\omega_o}s + \frac{1}{\omega_o^2}s^2)} \tag{4.15}$$

Die Lage des Polpaares (s_i, s_i^*) stellt sich dann in der komplexen s-Ebene gem. Abbildung 4.21 (rechte Bildhälfte) dar. Mit den eingeführten Größen läßt sich die kartesische Darstellung der komplexen Größe s_i auch anders darstellen:

$$s_i = \mathrm{Re}(s_i) + j\mathrm{Im}(s_i) = -d\omega_o + j\omega_e \tag{4.16}$$

Ein Übertragungssystem 2. Ordnung befindet sich für Dämpfungsgrade $0 \leq d < 1$ im Schwingfall, die Polstelle wandert dann in der s-Ebene je nach Wert der Variable d auf der gestrichelt eingezeichneten Ortskurve. Damit läßt sich die Eigenkreisfrequenz ω_e trigonometrisch berechnen zu

$$\omega_e = \omega_o\sqrt{1 - d^2} \tag{4.17}$$

In der Literatur wird die Formantfrequenz F_i unmittelbar aus dem Imaginärteil der komplexen Frequenz s_i, die zugehörige Formantbandbreite B_i aus dem Realteil der komplexen Frequenz s_i abgeleitet und es ergeben sich die bekannten Ausdrücke gem. Gleichung 4.11.

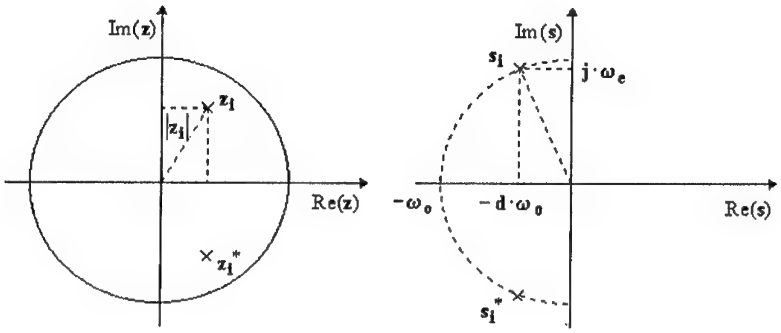

Abbildung 4.21.: Darstellung des komplexen Polpaares (z_i, z_i^*) der Teilsystemfunktion $H_i(z)$ in der komplexen z-Ebene (linke Bildhälfte) und des korrespondierenden komplexen Polpaares (s_i, s_i^*) in der komplexen s-Ebene (rechte Bildhälfte).

Ein Übertragungssystem 2. Ordnung kann aber nur dann eine tatsächliche spektrale Überhöhung zeigen, wenn die zugehörige Übertragungsfunktion $|H_{vok}(j\omega)|$ bei der entsprechenden Resonanzkreisfrequenz ω_r eine Amplitude erzeugt, die auch tatsächlich über der bei der Kreisfrequenz $\omega = 0$ liegt. Die Abbildung 4.22 soll diesen Zusammenhang veranschaulichen.
Gem. [Boe88] treten bei Übertragungssystemen 2. Ordnung Resonanzüberhöhungen erst dann auf, wenn für den Dämpfungsgrad d gilt:

$$d < \frac{1}{\sqrt{2}} \tag{4.18}$$

Der Ort der Resonanzüberhöhung liegt dann bei der Resonanzkreisfrequenz ω_r:

$$\omega_r = \omega_o \sqrt{1 - 2d^2} \tag{4.19}$$

und damit stets unterhalb der Eigenkreisfrequenz ω_e.

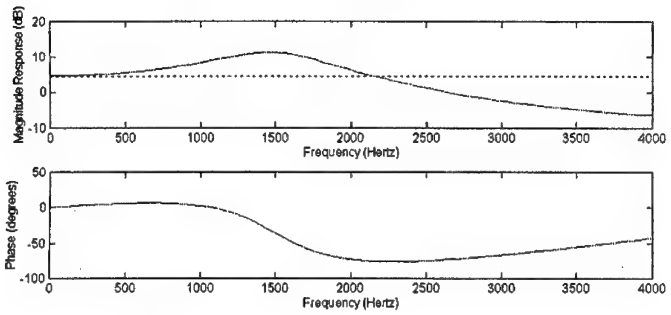

Abbildung 4.22.: Simulierter Amplitudengang eines Übertragungssystems 2.Ordnung mit einer Resonanzüberhöhung bei der Frequenz von 1500 Hz.

Nach Umrechung kann dieser Ort der Resonanzüberhöhung (und damit die exakte Formantfrequenz) auch direkt aus der Lage des Polpaares (z_i, z_i^*) abgeleitet werden:

$$F_r = \frac{\omega_r}{2\pi} = \frac{F_s}{2\pi}\sqrt{arc^2(z_i) - \ln^2|z_i|} \tag{4.20}$$

Die Gleichung 4.11 stellt also eine Näherung gegenüber Gleichung 4.20 dar, die exakten Werte für die Formantfrequenzen liegen stets unterhalb.

Um zu bewerten, wie sich diese Näherung auswirkt, ist in Abbildung 4.23 zunächst die geschätzte Formantfrequenz als Funktion gem. Gleichung 4.11 über der komplexen z-Ebene dargestellt.
Unabhängig vom Betrag einer Polstelle steigt bei Erhöhung der Phase die geschätzte Formantfrequenz linear an (Bewegung auf der dargestellten Oberfläche von oben gesehen im Uhrzeigersinn). Sie erreicht ihren Maximalwert der Nyquistfrequenz bei $arc(z) = \pi$.

Demgegenüber ist der nach Gleichung 4.20 exakt berechnete Verlauf der Resonanzfrequenz (also zugleich auch der tatsächlichen Formantfrequenz) in Abbildung 4.24 dargestellt.
Die jetzt entstehende Fläche zeigt besondere Eigenschaften:

- Die Resonanzfrequenz wächst nur am Rand (d.h. für $|z| \to 1$) direkt proportional zur Phase $arc(z)$ an.

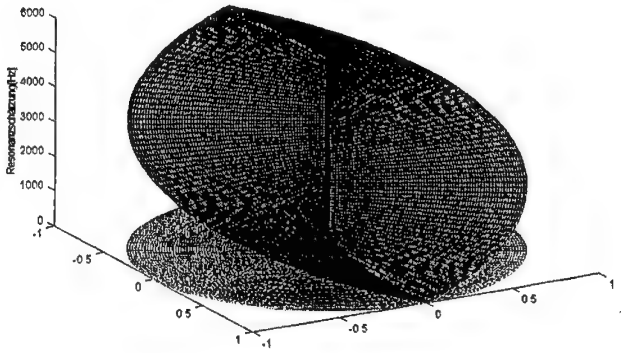

Abbildung 4.23.: Abhängigkeit der nach Gleichung 4.11 geschätzen Formantfrequenz von der Lage des korrespondierenden konjugiert komplexen Polpaares in der z-Ebene (die Imaginärachse der z-Ebene zeigt nach rechts). Bei konstantem Betrag und steigender Phase steigt auch die geschätze Formantfrequenz linear an.

- Je kleiner der Betrag $|z|$, desto stärker ist die Linearität des Zusammenhangs zwischen Phase $arc(z)$ und Resonanzfrequenz verletzt.
- Es gibt einen relevanten Bereich der z-Ebene, der *keine* Resonanzen zuläßt, auch wenn die Polstellen konjugiert komplex sind und damit prinzipiell resonanzfähig wären. Polstellen am Rand dieses Bereiches müssen zu sehr stark variierenden Resonanzfrequenzen führen, da sich hier die Neigung der Fläche abrupt ändert.

Die Differenz der beiden Kurven gem. Abbildung 4.25 stellt den absoluten Fehler grafisch dar, der bei der Näherung gem. Gleichung 4.11 gegenüber der exakten Lösung gem. Gleichung 4.20 entsteht. Der Fehler nimmt bei ungünstiger Pollage beträchtliche Werte an. Er sollte hier nicht vernachlässigt werden, da bei der häufig vorkommenden Hypernasalität sprechgestörter Patienten höhere Poldämpfungen auftreten, die diesen Effekt des Fehleranstiegs begünstigen.

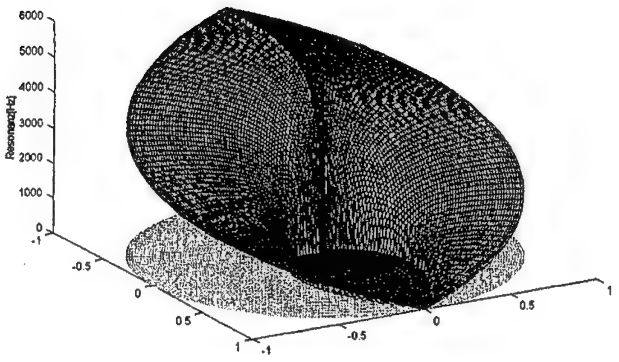

Abbildung 4.24.: Abhängigkeit der nach Gleichung 4.20 exakt berechneten Formantfrequenz von der Lage des korrespondierenden konjugiert komplexen Polpaares in der z-Ebene (die Imaginärachse der z-Ebene zeigt nach rechts). Gegenüber Abbildung 4.23 zeigen sich deutliche Abweichungen. Zusätzlich tritt hier ein relativ großer Bereich innerhalb der z-Ebene auf, der auch aus konjugiert komplexen Polpaaren gar keine Resonanzen ausbilden kann.

Bei der Realisierung von MODIAS wurde die exakte Lösung zur Berechnung der Formantfrequenzen berücksichtigt.

Die Selektion resonanzfähiger Polpaare (bei einem wie in MODIAS gewählten Prädiktorgrad K=16 gibt es 8 Polpaare) zur Gewinnung von Formantkandidaten für das nachgeordnete Formant-Tracking-Verfahren nutzt die besonderen Eigenschaften der Funktion gem. Gleichung 4.20 aus. Durch Umrechnung kann leicht ein Kriterium abgeleitet werden, welches diejenigen konjugiert komplexen Polpaare direkt in der z-Ebene ausschließt, die keine Resonanzüberhöhungen erzeugen können.

Ein konjugiert komplexes Polpaar (z_i, z_i^*) erzeugt nur dann eine sichtbare Resonanz

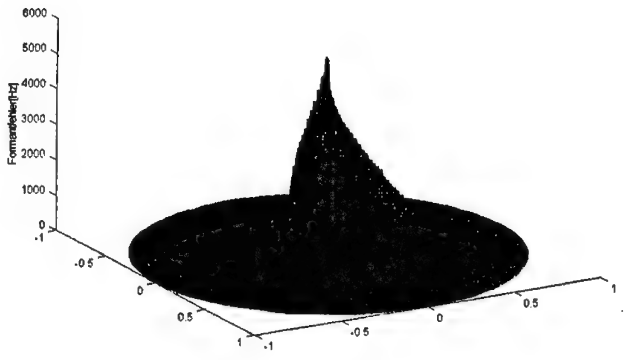

Abbildung 4.25.: Absoluter Fehler in der Berechnung der Formantfrequenz durch die üblich verwendete Näherung nach Gleichung 4.11 gegenüber der exakten Lösung nach Gleichung 4.20. Gerade die Polstellen mit höherer Dämpfung (also kleinerem Betrag $|z|$) führen in der Praxis häufig zu Störungen bis hin zur völligen Divergenz von nachgeordneten Formant-Trackingverfahren. Bei der Realisierung von MODIAS wurde dieser Effekt entsprechend berücksichtigt.

im Fourierspektrum (Formant), wenn Betrag und Phase in folgendem Zusammenhang stehen (Resonanzkriterium):

$$|z_i| > e^{-arc(z_i)} \tag{4.21}$$

Dieses Kriterium wurde bei der Realisierung des MODIAS-Moduls implementiert.

4.3.1.2. Formant-Tracking

Die vorangegangene Berechnung selektierte für jedes Signalsegment der Länge 256 diejenigen konjugiert komplexen Polpaare der Systemfunktion $H_{vok}(z)$, die auf Grund ihrer günstigen Eigenschaften spektrale Überhöhungen in der korrespondierenden Über-

tragungsfunktion $|H_{vok}(j\omega)|$ zu erzeugen vermögen. Aufgabe eines Formant-Tracking-Verfahrens ist es nun, aus dieser Teilmenge pro Segment genau die beiden Polpaare zu extrahieren, die die diagnostisch relevanten Formanten F1 und F2 repräsentieren und deren zeitlichen Verlauf von Segment zu Segment zu verfolgen. Dieses Problem ist nichttrivial und stellt die wohl größte Herausforderung an eine brauchbare Formantanalyse dar.
Bei der Realisierung des beschriebenen MODIAS-Moduls wird zunächst eine Polfrequenzmatrix Q gebildet. Jede Spalte j dieser Matrix korrespondiert dabei mit einem Signalsegment, die Zeilen i der Matrix beinhalten die noch verbliebenen Polfrequenzen des betreffenden Segmentes in frequenzmäßig aufsteigender Sortierung.
Im Idealfalle stellt nun die erste Zeile der Matrix bzw. $Q_{1,j}$ den zeitlichen Verlauf des Formanten F1 und die zweite Zeile $Q_{2,j}$ analog den Verlauf des Formanten F2 als Parameter-Zeit-Kontur dar. Bedingt u.a. durch variierende Phasenlagen des Sprachsignales im jeweiligen Segment ist eine solche starre Zuordnung zwischen Polfrequenzen und Formanten nicht immer möglich, da von einem Segment zum nächsten

- *echte* Formanten unvermittelt ausfallen (Formant-drop-outs) oder
- *unechte* Formanten sporadisch auftauchen, die im ungünstigen Falle frequenzmäßig unterhalb von F1 bzw. zwischen F1 und F2 liegen und damit fehlinterpretiert würden (Spectral-Shaping-Poles).

Normalerweise bedienen sich an dieser Stelle die publizierten Formant-Tracking-Verfahren eines Plausibilitätskriteriums durch Vergleich der avisierten Formantkandidaten mit tabellierten Normwerten für mittlere Formantlagen bei Normsprechern. Im Projekt MODIAS kann aber bedingt durch mannigfaltige Störungsprofile der sprechpathologischen Zielgruppe von keinerlei Normwerten mehr ausgegangen werden. Damit bleibt nur die Auswahl der *wahrscheinlichsten* Formantkandidaten aus der Polfrequenzmatrix selbst und die segmentweise Verfolgung dieser Kandidaten unter Zuhilfenahme heuristischer Plausibilitätskriterien, die sich im wesentlichen aus den physikalischen Massenträgheiten des menschlichen Vokaltraktes ableiten lassen.

Ableitung der Tracking-Startwerte für F1 und F2: Die Startwerte für das folgende Formant-Tracking werden aus einem heuristischen Ansatz gewonnen.

Für den Formant F1 wird aus der ersten Zeile der Polfrequenzmatrix bzw. $Q_{1,j}$ (Rohverlauf F1) der Median gebildet. Auf Grund seiner Berechnungsmodalität ist ein Median robust gegenüber Ausreißern und er wird durch solche - im Gegensatz zum Mittelwert - auch nicht verfälscht. Danach wird der Zeitpunkt (Spaltenindex j) gesucht, an dem der Rohverlauf F1 (Matrixzeile $Q_{1,j}$) den gerade ermittelten Median erstmals über- oder unterschreitet.
Der Spaltenindex j markiert damit den Startzeitpunkt, die zugehörige Polfrequenz markiert die Startfrequenz für das anschließende Formant-Tracking.
Die Ableitung des Startwerts für den Formant F2 wird analog durch Betrachtung des Rohverlaufes F2 (2. Zeile der Polfrequenzmatrix) gewonnen.

Tracking-Strategie: Das folgende Trackingprinzip reagiert auf die folgenden 4 denkbaren Störeinflüße:

- unvermittelter Ausfall des Formanten F1,
- unvermittelter Ausfall des Formanten F2,
- sporadisches Auftreten eines Spectral-Shaping-Poles unterhalb von F1,
- sporadisches Auftreten eines Spectral-Shaping-Poles zwischen F1 und F2.

Die Korrekturstrategie bei diesen 4 möglichen Störungsarten ist in Abbildung 4.26 grafisch veranschaulicht. Beim Übergang von einem Segment zum nächsten (Spalten!) wird stets die frequenzmäßig nächstgelegene Polfrequenz gewählt, die ein heuristisches Homogenitätskriterium (relative Änderung von einem Segment zum nächsten, also innerhalb von 5 ms maximal 5 Prozent) erfüllt. Dieser Vorgang wird in der Matrix in beiden Richtungen vom extrahierten Tracking-Startwert aus durchgeführt.

Glättung der Formantverläufe: Abschließend werden die beiden so gewonnenen Formantverläufe F1 und F2 einer 2-stufigen nichtlinearen Glättung nach Rabiner ([Rab75]) unterzogen. Dieses Glättungsverfahren wurde für das beschriebene Projekt herangezogen, da es die Vorteile einer linearen Glättung mit symmetrischem FIR-Filter (exakte Delaykompensation, Entfernung rauschartiger Störungscharakteristik möglich) und die Vorteile einer nichtlinearen Glättung mit Running-Median-Filter ungerader Länge (Entfernung

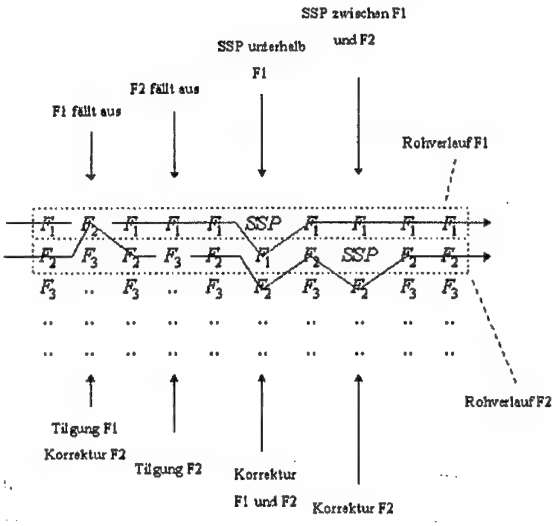

Abbildung 4.26.: Suchalgorithmus innerhalb der Polfrequenzmatrix Q zur Verfolgung der (echten) Formanten F1 und F2 bei den 4 möglichen Störungsarten durch Formant-Drop-Outs oder unerwünschter Spectral-Shaping-Poles (SSP).

von Ausreißern möglich, geglättete Werte sind stets in der Originalkontur enthalten) kombiniert. Das Blockschaltbild zum verwendeten Glättungsverfahren ist in Abbildung 4.27 dargestellt.

Die nichtlineare Glättung erfolgt hier durch zwei hintereinandergeschaltete Running-Median-Filter der Längen 5 und 3, die lineare Glättung durch FIR-Hamming-Filter der Länge 3.

Das vorgestellte Formantanalyse- und Formant-Tracking-Verfahren hat sich in zahlreichen Tests und praktischen Anwendungen bewährt. Dennoch sind Tracking-Fehler nicht auszuschließen, die besonders bei einem Versagen der erforderlichen Startwerte zu eklatanten Meßfehlern bzw. zur Divergenz des Suchalgorithmus führen können. Wie bereits beschrieben, können im konkreten Fall neurogener Sprechstörungen derart starke Abweichungen der Formantlagen von den Lagen gesunder Sprecher auftreten, daß der Anspruch an eine vollständig automatisierte Berechnung der akustischen Parameter nicht realistisch

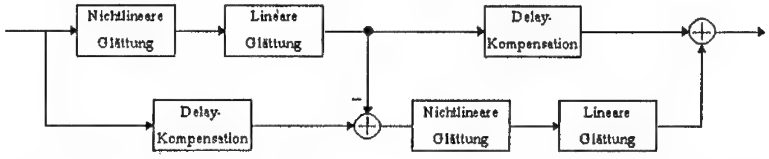

Abbildung 4.27.: Verwendetes Verfahren nach [Rab75] zur Glättung der zeitlichen Formantverläufe F1 und F2. Der Algorithmus kombiniert die Vorteile linearer und nichtlinearer Glättung und ist für Rohparameterverläufe mit Rauschstörungen und sporadischen Ausreißern konzipiert.

und eine fallweise Interaktion des Untersuchers unumgänglich erscheint. Daher wurde in das Untersuchungsprotokoll die Darstellung sogenannter LPC-Spektren integriert mit der Option, daß bei Verdacht auf startwertbedingte Trackingfehler entsprechend korrigierte Startwerte direkt in diesem LPC-Spektrum vorgegeben werden können. Sie bieten den Vorteil, die Formantfrequenzen wesentlich besser abzubilden als die Fourierspektren des Sprachsignales selbst.

Berechnung der LPC-Spektren: Die LPC-Spektren sind die korrespondierenden Übertragungsfunktionen zu den mittels linearer Prädiktion geschätzten Vokaltraktfiltern. Sie eliminieren systembedingt den Einfluß der Stimmbandschwingung und stellen daher eine geglättete Form des Sprachsignalspektrums dar. Die LPC-Spektren eignen sich daher besonders zur Visualisierung der Formanten, die im ursprünglichen Spektrum teilweise nur schwer erkennbar sind. Abbildung 4.28 zeigt das typische (rauhe) Leistungsdichtespektrum und das korrespondierende (glatte) LPC-Spektrum für den Vokal */i/*.
Das Vokaltraktfilter mit der Systemfunktion $H_{vok}(z)$ gem. Gleichung 4.6 ist im Zeitbereich allgemein durch eine Eingangsgröße $x(n)$ und eine Ausgangsgröße $y(n)$ darstellbar:

$$y(n) = x(n) + a_1 y(n-1) + a_2 y(n-2) + \ldots + a_K y(n-K) \tag{4.22}$$

Die Impulsantwort $h_{vok}(n)$ des Vokaltraktfilters auf den Einheitsimpuls $e(n) = [1, 0, 0, \ldots]$ ergibt sich durch Anwendung der Gleichung 4.22 zu:

$$h_{vok}(n) = [1, a_1, a_2, \ldots, a_K, 0, 0, 0, 0, 0 \ldots] \tag{4.23}$$

Damit stellen die Samples $2 \ldots K+1$ die Prädiktorkoeffizienten dar. MODIAS ermittelt die LPC-Spektren durch Fouriertransformation der einfach aufzustellenden Impulsant-

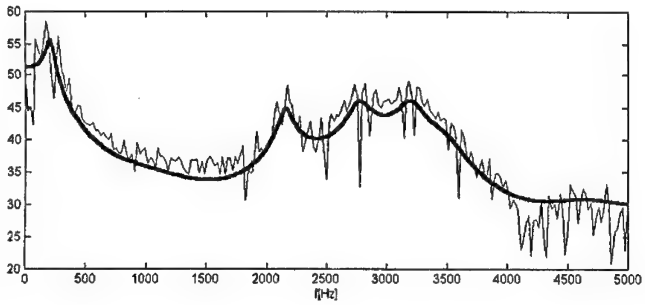

Abbildung 4.28.: Leistungsdichtespektrum (rauhe Kurve) und LPC-Spektrum (glatte Kurve) am Beispiel des Vokals */i/*. Während die Formanten im Leistungsdichtespektrum aufgrund des noch vorhandenen Einflusses der Stimmbandanregung schwer erkennbar sind, treten diese im LPC-Spektrum deutlich sichtbar hervor.

wort. Dabei wird die Impulsantwort zuvor mit Nullen auf die Länge 256 aufgefüllt.

4.3.1.3. Abgeleitete akustische Parameter zur Prüfung des Artikulationsraumes

Ausgangspunkt für alle abgeleiteten akustischen Parameter sind die ermittelten Formantverläufe F1 und F2. Dem Untersucher sollen ausschließlich diejenigen akustischen Parameter dargeboten werden, die eine hohe diagnostische Relevanz besitzen und ihn damit weder mit irrelevanter noch mit redundanter Information überfordern. Aus diagnostischer Sicht sind folgende akustische Informationen von Interesse:

- Beurteilung der mittleren Lagen der Formantpaare F1/F2 für jeden Vokal,
- Beurteilung der Stellungsänderung relevanter Artikulatoren durch Vergleich der Formanten verschiedener Vokale (z.B. F2-Absenkung von */i/* zu */y/* wegen Lippenrundung),
- Beurteilung der artikulatorischen Kontrastierfähigkeit und Prüfung des Artikulationsraumes durch Projektion der Formantpaare F1/F2 in die 2-dimensionalen F1/F2-Ebene (Formantkarte),

- Beurteilung der Lagestabilität relevanter Artikulatoren durch die Darstellung der zeitlichen Verläufe der Formanten F1 bzw. F2.

Im beschriebenen MODIAS-Modul werden hierzu aus den Formantverläufen durch Medianbildung die mittleren Formanten F1 bzw. F2 für jeden der betrachteten Vokale gewonnen. Die zeitlichen Verläufe wie auch die mittleren Formantlagen werden dem Untersucher in verschiedenen grafischen Darstellungsvarianten - wie bereits bei der Bedienung beschrieben - angeboten. Die Prüfung der artikulatorischen Kontrastierfähigkeit erfolgt über die Beurteilung der Fläche, die in der 2-dimensionalen F1/F2-Ebene von den Vokalen */i/*, */u/* und */a/* aufgespannt wird.

4.3.2. Verfahren zur Analyse von Stimmqualität und Stimmstabilität

4.3.2.1. Bestimmung des Grundfrequenzverlaufs und dessen Parametrierung

Die Basis jeder Stimmqualitäts- bzw. Stimmstabilitätsanalyse ist stets ein Algorithmus zur Extraktion des Grundfrequenzverlaufs. Gerade die Variabilität von einem Stimmritzenverschluß zum nächsten bzw. die Vollständigkeit des Stimmritzenverschlusses enthält dabei die diagnostisch relevante Information. Daher sind an diesen Algorithmus hohe Genauigkeitsanforderungen zu stellen.
Für eine akustische Analyse müssen somit die relevanten Schwankungen im Verlauf der Grundfrequenz zum einen abgebildet und durch geeignete Parameter beschrieben werden, zum anderen dürfen diese Effekte aber nicht durch Vorverarbeitungsschritte, wie z.B. Filterungen, verfälscht oder gar eliminiert werden. Zusätzlich gewinnt der Anspruch an ein geeignetes Verfahren weitere Komplexität, wenn mehrere Störungsmuster, wie in der Praxis häufig vorkommend, gleichzeitig auftreten. Dies darf nicht zur teilweisen oder gar vollständigen Divergenz des Algorithmus führen.
Die bekannten Verfahren zur Ermittlung des Grundfrequenzverlaufes lassen sich nach [Par99] in zwei Kategorien einteilen:

- Ereignis-Detektionsverfahren, die eine Schätzung der Grundfrequenz aufgrund partieller Ereignisse wie positiver oder negativer Peaks bzw. Nulldurchgänge im Oszillogramm vornehmen;
- Kurzzeit-Analyseverfahren, die eine Schätzung durch Mittelungen innerhalb eines einstellbaren Analysefensters gestatten.

Ein Vergleich der wichtigsten und zum Zeitpunkt der Erstellung der vorliegenden Arbeit aktuellen Verfahren ist in [Par99] publiziert. Im Rahmen dieses Vergleichs an normalen, gestörten und synthetischen Stimmen konnte gezeigt werden, daß die Leistungsfähigkeit sämtlicher Verfahren bei steigendem Stimmstörungsgrad sinkt oder ganz zum Erliegen kommt.
Keines der Verfahren liefert detailliertere Parameter zu den Stimmband-Schwingungsanomalien, z.B. für langsame Schwankungen der Grundfrequenz oder generelle Auf- bzw. Abwärtstrends in deren zeitlichem Verlauf.
Bei der Realisierung des beschriebenen MODIAS-Moduls wurde daher ein eigener

Algorithmus entwickelt, der die folgenden Störungsmuster von klinischer Relevanz im Ansatz berücksicht:

- Die Grundfrequenz kann während der Vokalisation einem Auf- oder Abwärtstrend unterliegen (Im Folgenden als *Pitch-Trend* bezeichnet).
- Die Grundfrequenz kann durch Störungen der Sprechatmung und Instabilitäten des Stimmbandtonus relativ langsam um den Trend schwanken (Im Folgenden als *Langsame Pitch-Variabilität* bezeichnet).
- Die Grundfrequenz kann relativ schnell, d.h. von Periode zu Periode variieren (Im Folgenden als *Jitter* bezeichnet).

Weiter konkretisiert bedeutet dies, daß die Parameter *Pitch-Trend*, *Pitch-Variabilität* und *Jitter* möglichst voneinander unabhängige Größen darstellen sollen.
Der hierzu entworfene Algorithmus detektiert im Zeitbereich sämtliche Schwingungsmaxima des Sprachsignales nach einem adaptiven Suchverfahren (Pitch-Perioden-Tracking) und unterzieht anschließend den so gewonnenen zeitlichen Grundfrequenzverlauf einer statistischen Analyse (Pitch-Verlaufsanalyse).

Pitch-Perioden-Tracking: Zur Detektion jedes einzelnen Schwingungsmaximums wurde hier ein Ereignisdetektionsverfahren entwickelt. Im Prinzip wird jedes Schwingungsmaximum des Sprachsignals mit einer (Pitch-)Marke versehen, deren zeitliche Abstände anschließend in Periodendauern bzw. Pitch-Frequenzen umgerechnet werden können. Entscheidend für die Konvergenz des Verfahrens sind drei Faktoren:

- Die erste Pitch-Marke zum Zeitpunkt t_1 muß auf einem tatsächlichen Schwingungsmaximum liegen (kein Artefakt!).
- Der zeitliche Abstand zur zweiten Pitch-Marke $(t_2 - t_1)$ muß durch eine gute Pitch-Periodendauer-Schätzung $\tilde{T}ini$ mit einer definierten, zeitlichen Unschärfe vorhergesagt werden können.
- Der Pitch darf während der Vokalisation des Probanden nur innerhalb definierter Grenzen schwanken.

Dem in digitaler Repräsentation vorliegenden Sprachsignal $s(n)$ werden zunächst nacheinander 3 Signalabschnitte (beginnend ab dem Einsatz der Vokalisation des Patienten) der Länge N=1024 entnommen. Durch das Kurzzeit-Verfahren der homomorphen bzw. cepstralen Analyse (z.B. beschrieben in [Hes83]) wird für jeden der 3 Abschnitte jeweils 1 Pitch-Grobschätzwert gewonnen; der Median dieser 3 Schätzungen bzw. dessen Kehrwert $\tilde{T}ini$ wird als Startwert für das folgende Ereignis-Detektionsverfahren herangezogen.

Die erste Pitch-Marke (Ereignis E_1 zum Zeitpunkt t_1) wird aus dem absoluten Maximum des Sprachsignals innerhalb des ersten Signalabschnitts der Länge. 1024 gewonnen. Das zweite Schwingungsmaximum wird jetzt, ausgehend vom Zeitpunkt t_1, in einem Suchfenster der Breite $[t_1 + 0.9\tilde{T}ini \ldots t_1 + 1.1\tilde{T}ini]$ wiederum durch Auswahl des absoluten Maximums bestimmt (Ereignis E_2 zum Zeitpunkt t_2).

Der ursprüngliche Schätzwert $\tilde{T}ini$ wird durch $Tper_1 = t_2 - t_1$ ersetzt und für die Bestimmung der folgenden Marke zum Zeitpunkt t_3 herangezogen, usw.
Bedingt durch die Festlegung der Suchfenster-Grenzen kann dieser Algorithmus einer relativen Periodendauer-Schwankung von maximal 10 Prozent (von einer Schwingung zur nächsten) nach oben oder unten folgen.

MODIAS erstellt für jede Vokalaufzeichnung einen Ereignisvektor E, dessen Spalten mit den Einzelschwingungen $i = 1 \ldots K$ korrespondieren. Jede Spalte beinhaltet dabei den detektierten Beobachtungszeitpunkt t_i des jeweiligen Ereignisses bzw. Schwingungsmaximums:

$$E = \begin{pmatrix} t_1 \\ t_2 \\ \cdot\cdot \\ t_K \end{pmatrix} \tag{4.24}$$

Die Periodendauern der Einzelschwingungen ergeben sich durch Differenzbildung zwischen jeweils zwei aufeinanderfolgenden Beobachtungszeitpunkten, bzw. $Tper_i = t_{i+1} - t_i$ für $i = 1 \ldots K-1$. Damit entsteht aus den $K-1$ Periodendauern ein Vektor $Tper$ gemäß

folgender Berechnungsweise:

$$Tper = \begin{pmatrix} Tper_1 \\ Tper_2 \\ \cdots\cdots\cdots \\ Tper_{K-1} \end{pmatrix} = \begin{pmatrix} t_2 - t_1 \\ t_3 - t_2 \\ \cdots\cdots\cdots \\ t_K - t_{K-1} \end{pmatrix} \tag{4.25}$$

Durch einfache Kehrwertbildung und Multiplikation mit der Abtastrate F_s entsteht ein Vektor $Froh$, der den Rohverlauf der Grundfrequenz (Pitch) sämtlicher Einzelschwingungen repräsentiert. Er wird wie folgt gebildet::

$$Froh = \begin{pmatrix} Froh_1 \\ Froh_2 \\ \cdots\cdots\cdots \\ Froh_{K-1} \end{pmatrix} = F_s \begin{pmatrix} \frac{1}{t_2 - t_1} \\ \frac{1}{t_3 - t_2} \\ \cdots\cdots\cdots \\ \frac{1}{t_K - t_{K-1}} \end{pmatrix} \tag{4.26}$$

Dieser Roh-Pitchverlauf wird nun im Rahmen einer statistischen Verlaufsanalyse eingehender betrachtet.

Pitch-Verlaufsanalyse: Wie bereits beschrieben, werden für die weitere Bearbeitung 3 Arten der Instabilität im Pitchverlauf definiert bzw. unterschieden, das sind *Pitch-Trend*, *Pitch-Variabilität* und *Jitter*. Eine Veranschaulichung hierzu bietet Abbildung 4.29.

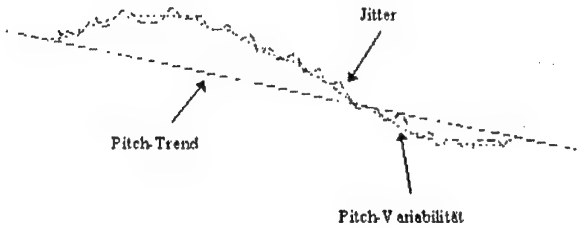

Abbildung 4.29.: MODIAS unterscheidet 3 verschiedene Arten von Instabilitäten im Pitchverlauf: *Pitch-Trend*, *Pitch-Variabilität* und *Jitter*.

Hierzu wird der Roh-Pitchverlauf $Froh$ in einem ersten Bearbeitungsschritt nach einem zweistufigen Verfahren geglättet (Blockschaltbild und Dimensionierung identisch zu Abbildung 4.27 auf Seite 86). Nach Glättung liegt der Pitchverlauf F vor, der z.B.

über einer rekonstruierten Zeitachse für alle realisierten Vokale wie auf den MODIAS-Auswerteblättern gem. Abbildung 4.11 dargestellt werden kann.
In einem zweiten Schritt wird aus dem gewonnenen geglätteten Pitchverlauf F durch lineare Regression eine Trendgerade $Ftre$ errechnet (Verfahren z.B. in [Bro79], siehe auch Abbildung 4.29).
Durch Normierung auf den mittleren Pitch entsteht der Vektor $\dot{F}tre$, der den relativen Trend des Pitchverlaufs beschreibt:

$$\dot{F}tre = 100 \frac{Ftre - \frac{1}{K-1} \sum_{i=1}^{K-1} F_i}{\frac{1}{K-1} \sum_{i=1}^{K-1} F_i} \tag{4.27}$$

Durch Differenzbildung und Normierung entsteht ein neuer Vektor $\dot{F}var$, der die relative (trendbereinigte) Schwankung des (geglätteten) Pitchverlaufs um seinen linearen Trend beschreibt:

$$\dot{F}var = \begin{pmatrix} \dot{F}var_1 \\ \dot{F}var_2 \\ \cdots\cdots\cdots \\ \dot{F}var_{K-1} \end{pmatrix} = 100 \begin{pmatrix} \frac{F_1 - Ftre_1}{Ftre_1} \\ \frac{F_2 - Ftre_2}{Ftre_2} \\ \cdots\cdots\cdots\cdots \\ \frac{F_{K-1} - Ftre_{K-1}}{Ftre_{K-1}} \end{pmatrix} \tag{4.28}$$

Im dritten und letzten Bearbeitungsschritt wird vom Roh-Pitchverlauf $Froh$ der geglättete Pitchverlauf F subtrahiert und auf diesen normiert. Es entsteht der Vektor $\dot{F}jit$, der die relative Schwankung des Roh-Pitchverlaufs um den (geglätteten) Pitchverlauf beschreibt:

$$\dot{F}jit = \begin{pmatrix} \dot{F}jit_1 \\ \dot{F}jit_2 \\ \cdots\cdots\cdots \\ \dot{F}jit_{K-1} \end{pmatrix} = 100 \begin{pmatrix} \frac{Froh_1 - F_1}{F_1} \\ \frac{Froh_2 - F_2}{F_2} \\ \cdots\cdots\cdots\cdots \\ \frac{Froh_{K-1} - F_{K-1}}{F_{K-1}} \end{pmatrix} \tag{4.29}$$

Die beiden Verläufe $\dot{F}var$ und $\dot{F}jit$ sind jetzt über das statistische Maß der Standardabweichung parametrierbar. MODIAS stellt die folgenden akustischen Parameter zur Verfügung:

- JITTER: Standardabweichung des normierten Jitters $\dot{F}jit_i$. Der Parameter beschreibt die relative Schwankung des Roh-Pitchverlaufs um den geglätteten Pitchverlauf und trägt die Dimension *Prozent*.

- PITCH-VARIABILITÄT: Standardabweichung der normierten Pitch-Variabilität $\dot{F}var_i$. Der Parameter beschreibt die relative Schwankung des geglätteten Pitchverlaufes um seinen linearen Trend und trägt die Dimension *Prozent*.

- PITCH-TREND: Steigung der auf den mittleren Pitch normierten Trendgeraden $\dot{F}tre$. Der Parameter beschreibt den relativen linearen Abfall/Anstieg des Pitchverlaufs und trägt die Dimension *Prozent pro Sekunde*.

4.3.2.2. Bestimmung des Lautheitsverlaufs und dessen Parametrierung

Während sich die vorangegangenen Betrachtungen vorwiegend auf die zeitlichen Aspekte der Stimmbandschwingung bzw. deren Periodendauer konzentrierten, liegt in diesem Abschnitt der Schwerpunkt auf der Betrachtung der phonatorischen Intensität dieser Schwingungen, also ihrer Amplitude. Hier wäre prinzipiell ein analoges Verfahren möglich, da der beschiebene Pitch-Tracking-Algorithmus bereits die relevanten Schwingungsmaxima im Oszillogramm detektiert und hier anstelle ihrer zeitlichen Lagen nur die entsprechenden Amplituden auszuwerten wären. Die schnellen Variationen der Intensität, in Analogie zum (Pitch-)Jitter üblicherweise als (Amplituden-)Shimmer bezeichnet, liefern gem. einschlägiger Literatur eine diagnostische Aussage, die sehr stark mit der des Jitters korreliert, also einer perzeptiv empfundenen Stimmrauhigkeit. Aufgrund dieser Informationsredundanz wurde auf die Ermittlung des Shimmers verzichtet.
Damit entfällt hier letztlich der hohe Anspruch eines Ereignisdetektionsverfahrens, da nicht mehr die Amplitude jeder Einzelschwingung bestimmt werden muß. Bei der MODIAS-Realisierung wurde daher für die weitere Betrachtung als Intensitätsmaß die Empfindungsgröße der (psychoakustischen) Lautheit herangezogen. Sie wird durch ein Kurzzeit-Analyseverfahren gewonnen und berücksichtigt auch die psychoakustischen Effekte der Sprachwahrnehmung.
Die komplexe Berechnung der zeitlichen Lautheitsverläufe ist in 5.3.1 auf den Seiten 142 ff. im Detail dargestellt und wird hier nicht weiter ausgeführt. Die Verläufe werden dem Untersucher im vorliegenden Modul STIMME UND VOKALARTIKULATION im Laufe des Untersuchungsprotokolls als Hilfmittel bei der Segmentierung geeigneter Signalabschnitte grafisch dargeboten (siehe Abbildung 4.7 auf Seite 54).

Folgende Störungsmuster sind im Verlauf der Lautheit diagnostisch relevant und waren daher bei der Realisierung entsprechend zu berücksichtigen:

- Die Lautheit kann während der Vokalisation einem Auf- oder Abwärtstrend unterliegen (Im Folgenden als *Lautheits-Trend* bezeichnet).
- Die Lautheit kann durch Störungen der Sprechatmung zeitlichen Schwankungen unterliegen (Im Folgenden als *Lautheits-Variabilität* bezeichnet).

Wie schon bei der Analyse des Pitchverlaufs bedeutet dies auch hier, daß die beiden vorgestellten Größen möglichst voneinander unabhängig reproduzierbar sein sollen.

Parametrische Repräsentation des Lautheitsverlaufs: Der Lautheitsverlauf entsteht, wie bereits erwähnt, durch ein Kurzzeit-Analyseverfahren. Hierzu wird das in zeitdiskreter Form vorliegende Sprachsignal $s(n)$ im Rahmen der Vorverarbeitung in zeitlich überlappende Signalsegmente zu je 512 Samples segmentiert. Pro Segment entsteht damit jeweils 1 diskreter Lautheitswert. Die Überlappung wurde derart gewählt, daß in einem festen Zeitraster von 10 ms neue Segmente beginnen bzw. Lautheitswerte entstehen.
MODIAS erstellt somit für jede Vokalaufzeichnung einen Vektor $Nroh$, dessen Spalten mit den Segmenten $i = 1 \ldots K$ korrespondieren. Er repräsentiert den (Roh-)Verlauf der Lautheit aller K Segmente:

$$Nroh = \begin{pmatrix} Nroh_1 \\ Nroh_2 \\ \ldots\ldots \\ Nroh_K \end{pmatrix} \tag{4.30}$$

Dieser Roh-Lautheitsverlauf wird nun bezüglich seiner zeitlichen Variabilität eingehender betrachtet.

Lautheits-Verlaufsanalyse: Wie bereits beschrieben, werden für die weitere Bearbeitung nur 2 Arten der Instabilität im Lautheitsverlauf definiert bzw. unterschieden, das sind *Lautheits-Trend* und *Lautheits-Variabilität*. Eine Veranschaulichung hierzu bietet Abbildung 4.30.

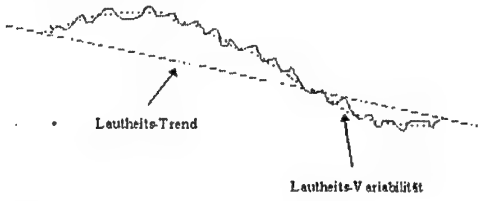

Abbildung 4.30.: MODIAS unterscheidet 2 verschiedene Arten von Instabilitäten im Lautheitsverlauf: *Lautheits-Trend* und *Lautheits-Variabilität.*

Hierzu wird der Roh-Lautheitsverlauf $Nroh$ in einem ersten Bearbeitungsschritt nach einem zweistufigen Verfahren geglättet (Blockschaltbild und Dimensionierung identisch zu Abbildung 4.27 auf Seite 86). Nach Glättung liegt der Lautheitsverlauf N vor (Darstellung des Lautheitsverlaufs N über einer rekonstruierten Zeitachse für alle realisierten Vokale innerhalb der MODIAS-Segmentierungsoberfläche gem. Abbildung 4.7.).
In einem zweiten Schritt wird aus dem gewonnenen (geglätteten) Lautheitsverlauf N durch lineare Regression eine Trendgerade $Ntre$ errechnet (Verfahren z.B. in [Bro79], siehe auch Abbildung 4.30).
Durch Normierung auf die mittlere Lautheit entsteht der Vektor $\dot{N}tre$, der den relativen Trend des Lautheitsverlaufs beschreibt:

$$\dot{N}tre = 100 \frac{Ntre - \frac{1}{K}\sum_{i=1}^{K} N_i}{\frac{1}{K}\sum_{i=1}^{K} N_i} \tag{4.31}$$

Durch Differenzbildung und Normierung entsteht ein neuer Vektor $\dot{N}var$, der die relative (trendbereinigte) Schwankung des (geglätteten) Lautheitsverlaufs um seinen linearen Trend beschreibt:

$$\dot{N}var = \begin{pmatrix} \dot{N}var_1 \\ \dot{N}var_2 \\ \cdots\cdots \\ \dot{N}var_K \end{pmatrix} = 100 \begin{pmatrix} \frac{N_1 - Ntre_1}{Ntre_1} \\ \frac{N_2 - Ntre_2}{Ntre_2} \\ \cdots\cdots\cdots \\ \frac{N_K - Ntre_K}{Ntre_K} \end{pmatrix} \tag{4.32}$$

Der Verlauf $\dot{N}var$ ist jetzt über das statistische Maß der Standardabweichung parametrierbar. MODIAS stellt die folgenden akustischen Parameter zur Verfügung:

- LAUTHEITS-VARIABILITÄT: Standardabweichung der normierten Lautheits-Variabilität $\dot{N}var_i$. Der Parameter beschreibt die relative Schwankung des Lautheitsverlaufes um seinen linearen Trend und trägt die Dimension *Prozent*.
- LAUTHEITS-TREND: Steigung der auf die mittlere Lautheit normierten Trendgeraden $\dot{N}tre$. Der Parameter beschreibt den relativen linearen Abfall/Anstieg des Lautheitsverlaufs und trägt die Dimension *Prozent*.

4.3.2.3. Bestimmung des CPPS-Verlaufs und dessen Parametrierung

Im Rahmen früherer Arbeiten haben sich zur akustischen Prädiktion perzeptiver Behauchtheitsratings drei verschiedene Charakteristika des Sprachsignales als wesentlich herausgestellt ([Hil96]):

- Relative Amplitude der ersten Harmonischen im Fourier-Leistungsdichtespektrum (das ist die charakteristische Spektrallinie bei der Stimmbandgrundfrequenz),
- Grad der Periodizität des zeitlichen Sprachsignales,
- Spectral Tilt (Abfall des Fourier-Leistungsdichtespektrums zu hohen Frequenzen).

Eine 1996 von HILLENBRAND und HOUDE durchgeführte Untersuchung an 20 pathologischen und 5 gesunden Sprechern vergleicht mehrere diesbezügliche Verfahren in der Korrelation mit perzeptiven Daten bei gehaltenen Vokalen. Dabei konnte eine deutliche Überlegenheit des komplexen CPPS-Parameters (Abkürzung für *Cepstral Peak Prominence Smoothed*) gezeigt werden bei einer Korrelation von über 96 Prozent mit den perzeptiven Ratings geschulter Hörer ([Hil96]). Zusätzlich bietet diese Methode den Vorteil, ohne Bedienerinteraktion oder parallele Algorithmenkontrolle sehr stabile Ergebnisse zu liefern - ein besonders wesentlicher Gesichtspunkt im Rahmen der beschriebenen MODIAS-Realisierung.
In der entsprechenden Untersuchung gem. [Hil96] konnte der CPPS-Parameter nur für ausgewählte, handsegmentierte und vermeintlich repräsentative Signalabschnitte kurzer Dauer ermittelt werden.

Bei der Realisierung des beschriebenen MODIAS-Moduls wurde eine automatische

Berechnung dieses CPPS-Parameters (100 Werte pro Sekunde) ohne manuelle Interaktion implementiert. Der Untersucher hat somit die Möglichkeit, auch den zeitlichen CPPS-Verlauf über die Dauer der Vokalproduktion hinweg zu beobachten.

HILLENBRAND und HOUDE berichten aufgrund ihrer experimentellen Validierung nur von der hohen Korrelation zwischen CPPS und perzeptiver Behauchtheit, äußern hingegen aber auch die Vermutung, daß sich dieser Parameter ebenfalls zur akustischen Prädiktion der perzeptiven Rauhigkeit eignen könnte. Die im Rahmen der MODIAS-Realisierung vorgestellte Überprüfung anhand eines speziell zusammengestellten synthetischen Sprachmaterials stützt diese Einschätzung.

Besondere Eigenschaften des Cepstrums: Der zu Grunde liegende Algorithmus zur Berechnung der CPPS bedient sich der homomorphen Analyse durch Berechnung des Cepstrums (z.B. beschrieben in [Hes83]). Das Cepstrum (der Begriff *Cepstrum* ist eine Wortschöpfung durch Umkehrung der ersten Silbe des Wortes *Spectrum*) ist eine dem Fourier-Spektrum ähnliche Signalrepräsentation.

Das herkömmliche Fourier-Spektrum, bzw. genauer die komplexe, spektrale Amplitudendichte $S(e^{j\omega})$ eines periodischen Zeitsignals $s(t)$ entsteht durch die in Abbildung 4.31 dargestellte signalverarbeitende Struktur. Zunächst wird ein geeigneter Signalausschnitt durch Multiplikation mit einem Fenster extrahiert, der dann der eigentlichen Fourier-Transformation (Fourier-Operator $FFT\{\}$) unterzogen wird.

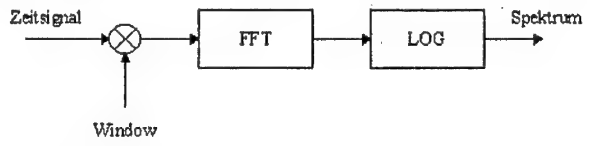

Abbildung 4.31.: Signalverarbeitende Struktur zur Berechnung des Fourier-Spektrums. Die Analyse/Interpretation des Signales s(t) erfolgt hier anhand der Eigenschaften der spektralen Signalrepräsentation.

Die abschließende Logarithmierung (Logarithmus-Operator $LOG\{\}$) dient unter ande-

rem der praktikablen Skalierung des resultierenden Spektrums.

Für die weitere Betrachtung wird nochmals das Quelle-Filter-Modell gem. Abbildung 4.17 herangezogen. Das Sprachsignal s(t) entsteht nach diesem Modell aus dem Anregungssignal e(t) durch Passieren des Vokaltraktfilters und damit durch Faltung (Operator $\otimes$) mit dessen Impulsantwort $h_{vok}(t)$:

$$s(t) = e(t) \otimes h_{vok}(t) \tag{4.33}$$

Das bedeutet, daß durch Anwendung der Struktur gem. Abbildung 4.31 das korrespondierende Fourier-Spektrum $S(e^{j\omega})$ sich auch als Produkt darstellen läßt:

$$S(e^{j\omega}) = E(e^{j\omega}) \cdot H(e^{j\omega}) \tag{4.34}$$

Die Anwendung des Logarithmus-Operators $LOG\{\}$ zeigt hier einen zweiten Vorteil, nämlich die Umwandlung eines Produkts in eine Summe und damit eine leichte Trennbarkeit der beiden Einzel-Fourier-Spektren von originärer Anregung (hier: Stimme) und Filterwirkung (hier: Artikulation):

$$LOG\{S(e^{j\omega})\} = LOG\{E(e^{j\omega}) \cdot H(e^{j\omega})\} = LOG\{E(e^{j\omega})\} + LOG\{H(e^{j\omega})\} \tag{4.35}$$

Ein typisches (logarithmiertes) Vokal-Spektrum ist in Abbildung 4.32 dargestellt. Das periodische Anregungssignal $e(t)$ korrespondiert mit periodisch wiederkehrenden Spektrallinien (harmonische Vielfache der Stimmbandgrundfrequenz F bzw. Pitch) und somit einem diskreten Linienspektrum (Feinstruktur), die Impulsantwort des Vokaltraktfilters mit einem additiv verknüpften, kontinuierlichen Spektrum (Grobstruktur).
Das Cepstrum $s(\tau)$ eines periodischen Zeitsignals $s(t)$ entsteht demgegenüber durch die in Abbildung 4.33 dargestellte Struktur. Wie zuvor wird ein geeigneter Signalausschnitt durch Anwendung eines Fensters extrahiert, der sowohl einer Fourier-Transformation als auch einer Logarithmierung unterzogen wird.
Anschließend erfolgt hier jedoch die Fourier-Rücktransformation des logarithmierten Spektrums (Inverser Fourier-Operator $IFFT\{\}$). Durch die dazwischenliegende Logarithmus-Operation ist das resultierende Cepstrum nicht mehr eine Funktion der Zeit, sondern der neuen Variablen Quefrency (der Begriff *Quefrency* ist eine Wortschöpfung durch eine Silbenumstellung des Wortes *Frequency*).
Durch Anwendung der Gleichung 4.35, sowie des Superpositionssatzes der Fourier-

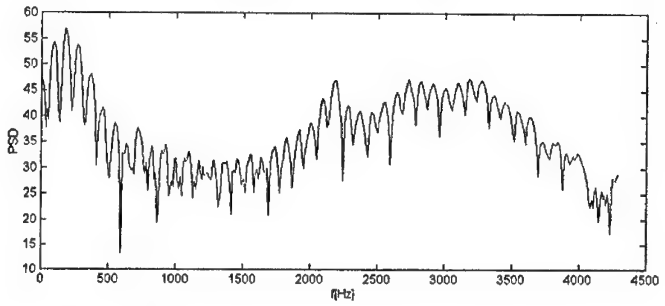

Abbildung 4.32.: Typisches, logarithmiertes Fourier-Spektrum eines Vokals. Es setzt sich additiv aus einem Linienspektrum (Einfluß der periodischen Stimmbandanregung) und einem kontinuierlichem Spektrum (Einfluß des Vokaltraktfilters) zusammen.

Transformation läßt sich dieser Zusammenhang wie folgt darstellen:

$$IFFT\{LOG\{S(e^{j\omega})\}\} = \overbrace{\underbrace{IFFT\{LOG\{E(e^{j\omega})\}\}}_{e(\tau)} + \underbrace{IFFT\{LOG\{H(e^{j\omega})\}}_{h(\tau)}}^{s(\tau)} \quad (4.36)$$

Nach Anwendung des Logarithmus-Operators und der Fourier-Rücktransformation ist also ein Cepstrum $s(\tau)$ entstanden. Es setzt sich additiv aus zwei Einzelcepstren zusammen, die mit der Anregung (hier: $e(\tau)$) und der Filterwirkung (hier: $h(\tau)$) korrespondieren.
Das zum Fourier-Spektrum gem. Abbildung 4.32 gehörende Vokal-Cepstrum ist in Abbildung 4.34 dargestellt.
Das periodische Anregungssignal (Stimme) korrespondiert mit einer einzigen diskreten Cepstrallinie bei der Quefrency $1/F$ bzw. dem Kehrwert der Stimmbandgrundfrequenz (hier: 11ms). Der Einfluß des Vokaltraktes (Artikulation) ist an den Rand des Cepstrums zu niedrigen Quefrencies hin gedrängt.
Diese besondere Eigenschaft des Cepstrums, eine solche diskrete Cepstrallinie auszubilden, wird bei der Ableitung des CPPS-Parameters genutzt.

Berechnung des CPPS-Verlaufs: HILLENBRAND und HOUDE haben in ihrer Untersuchung gezeigt, daß die Amplitude dieser dominanten Komponente $s(\tau_p) = s(1/F)$ so-

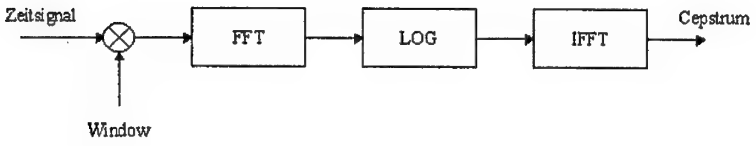

Abbildung 4.33.: Signalverarbeitende Struktur zur Berechnung des Cepstrums. Die Analyse/Interpretation des Signales s(t) erfolgt hier anhand der Eigenschaften der cepstralen Signalrepräsentation.

wohl von der momentanen Signalleistung als auch vom Grad der Periodizität der Stimmbandschwingung abhängt.
Um den Einfluß der momentanen Signalleistung zu eliminieren, wird gem. ([Hil96]) eine Regressionsgerade bzw. ein linearer Trend aus dem Cepstrum im Quefrency-Bereich $\tau = 33 \ldots 167$ ms errechnet. Das CPPS-Maß beschreibt die Amplitudendifferenz in dB, um die die Cepstrallinie über diese Regressionsgerade bei der Quefrency $\tau_p = 1/F$ herausragt. Dieser Zusammenhang ist in Abbildung 4.35 grafisch dargestellt.
Der CPPS-Verlauf erwächst aus der Analyse segmentweise errechneter Cepstren gem. Abbildung 4.33. Hierzu wird das in zeitdiskreter Form vorliegende Sprachsignal im Rahmen der Vorverarbeitung in zeitlich überlappende Signalsegmente zu je 1024 Samples segmentiert, pro Segment entsteht damit jeweils 1 diskreter CPPS-Wert. Die Überlappung wurde derart gewählt, daß in einem festen Zeitraster von 2 ms neue Segmente beginnen bzw. CPPS-Werte entstehen. Als Fenster gem. Abbildung 4.33 wird in [Hil96] eine Hamming-Form vorgeschlagen (siehe Abbildung 4.19), ebenso eine Fast-Fourier-Transformation bzw. -Rücktransformation der Länge 1024.
Die Einzelcepstren werden vor Extraktion der Amplitude der jeweils dominanten Cepstrallinie in zwei Stufen geglättet:

- Stufe 1: Mittelung über die Zeit. In Abständen von 10ms wird dabei jeweils ein vorgeglättetes Cepstrum des Segmentes i durch Mittelung über die 9 Cepstren der Segmente $i-4 \ldots i+4$ gebildet (9-Frame-Cepstral-Time-Average).

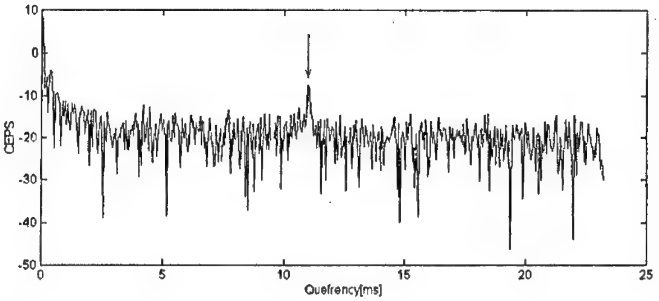

Abbildung 4.34.: Typisches Cepstrum eines Vokals. Es setzt sich additiv aus dem Cepstrum des Anregungssignals (dominante Cepstrallinie beim Kehrwert des Pitchs bzw. $\tau_p = 1/F$) und dem Cepstrum des Vokaltraktfilters (Bereich $\tau << \tau_p$) zusammen.

- Stufe 2: Filterung der vorgeglätteten Cepstren durch ein 3-Punkt-Running-Median-Filter.

Nach Glättung und Extraktion stehen also CPPS-Werte im Abstand von 10 ms zur Verfügung.
MODIAS erstellt für jede Vokalaufzeichnung einen Vektor $Croh$, dessen Spalten mit den Segmenten $i = 1 \ldots K$ korrespondieren. Er repräsentiert den (Roh-)Verlauf des CPPS-Parameters aller K Segmente:

$$Croh = \begin{pmatrix} Croh_1 \\ Croh_2 \\ \ldots\ldots \\ Croh_K \end{pmatrix} \tag{4.37}$$

Dieser Roh-CPPS-Verlauf wird nun bezüglich seiner zeitlichen Variabilität eingehender betrachtet.

Statistische CPPS-Verlaufsanalyse: Für die weitere Bearbeitung werden 2 Arten der Instabilität im CPPS-Verlauf definiert bzw. unterschieden, das ist der *CPPS-Trend* und

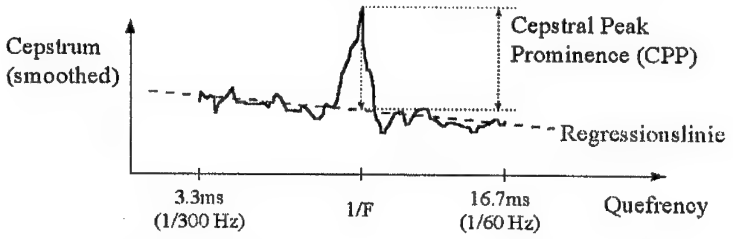

Abbildung 4.35.: Der CPPS-Parameter beschreibt die Differenz zwischen der Amplitude der dominanten Cepstrallinie bei der Quefrency $\tau_p = 1/F$ und der ermittelten Regressionsgeraden in dB. Diese Art der Berechnung eliminiert den Einfluß der momentanen Signalleistung.

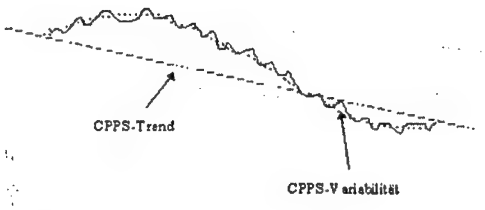

Abbildung 4.36.: MODIAS unterscheidet 2 verschiedene Arten von Instabilitäten im CPPS-Verlauf: *CPPS-Trend* und *CPPS-Variabilität.*

die *CPPS-Variabilität.* Eine grafische Veranschaulichung hierzu bietet Abbildung 4.36. Der Roh-CPPS-Verlauf *Croh* wird in einem ersten Bearbeitungsschritt nach einem zweistufigen Verfahren geglättet (Blockschaltbild und Dimensionierung identisch zu Abbildung 4.27 auf Seite 86). Nach Glättung liegt der CPPS-Verlauf C vor (Darstellung des CPPS-Verlaufs C über einer rekonstruierten Zeitachse für alle realisierten Vokale auf den MODIAS-Auswerteblättern gem. Abbildung 4.12).
In einem zweiten Schritt wird aus dem gewonnenen (geglätteten) CPPS-Verlauf C durch lineare Regression eine Trendgerade *Ctre* errechnet (Verfahren z.B. in [Bro79]).

Durch Normierung auf die mittlere CPPS entsteht der Vektor $\dot{C}tre$, der den rela-

tiven Trend des CPPS-Verlaufs beschreibt:

$$\dot{C}tre = 100 \frac{Ctre - \frac{1}{K}\sum_{i=1}^{K} C_i}{\frac{1}{K}\sum_{i=1}^{K} C_i} \tag{4.38}$$

Durch Differenzbildung und Normierung entsteht ein neuer Vektor $\dot{C}var$, der die relative (trendbereinigte) Schwankung des (geglätteten) CPPS-Verlaufs um seinen linearen Trend beschreibt:

$$\dot{C}var = \begin{pmatrix} \dot{C}var_1 \\ \dot{C}var_2 \\ \cdots\cdots \\ \dot{C}var_K \end{pmatrix} = 100 \begin{pmatrix} \frac{C_1 - Ctre_1}{Ctre_1} \\ \frac{C_2 - Ctre_2}{Ctre_2} \\ \cdots\cdots\cdots \\ \frac{C_K - Ctre_K}{Ctre_K} \end{pmatrix} \tag{4.39}$$

Der Verlauf $\dot{C}var$ ist jetzt über das statistische Maß der Standardabweichung parametrierbar. MODIAS stellt die folgenden akustischen Parameter zur Verfügung:

- CPPS-VARIABILITÄT: Standardabweichung der normierten CPPS-Variabilität $\dot{C}var_i$. Der Parameter beschreibt die relative Schwankung des CPPS-Verlaufes um seinen linearen Trend und trägt die Dimension *Prozent*.
- CPPS-TREND: Steigung der auf die mittlere CPPS normierten Trendgeraden $\dot{C}tre$. Der Parameter beschreibt den relativen linearen Abfall/Anstieg des CPPS-Verlaufs und trägt die Dimension *Prozent*.

4.4. Validitätsprüfung der Komponente STIMME

Das Hauptziel akustischer Stimmanalysen ist es, mit möglichst hoher Güte das Urteil eines geschulten Experten vorherzusagen und damit einen reproduzierbaren und objektiven Zusammenhang zwischen akustischer und auditiver Information herzustellen. Die statistische Validität des Modul-Anteils STIMME wurde im Rahmen der Enwicklung in zwei unabhängigen Schritten eingehend geprüft.

4.4.1. Validität der gewählten Untersuchungsmethode

In diesem Abschnitt wird zunächst die Wahl der Untersuchungsmethode (isolierte Vokale) begründet. Danach wird die Frage untersucht, ob prinzipiell von den hieraus gewonnenen Stimmparametern auf die entsprechenden Parameter im natürlichen Redefluß geschlossen werden darf. Dabei ist zu berücksichtigen, daß die gewählte Aufgabenstellung isolierter Vokale artifiziell ist und einer Fixierung des Kehlkopfs über einen vergleichsweise langen Zeitraum bedarf. Eine derartige Konfiguration tritt im natürlichen Redefluß grundsätzlich nicht auf.
Die Entscheidung für die gewählte Aufgabenstellung isolierter Vokale hatte folgende Gründe:

- Die Aufgabe kann auch von Patienten mit sehr schweren neurologischen Störungen noch hinreichend erfüllt werden.
- Pitch-Tracking-Algorithmen arbeiten hier in der Regel wesentlich stabiler als im natürlichen Redefluß.
- Es ist keine Segmentierung von Vokalabschnitten erforderlich, daher kann auf eine stimmhaft/stimmlos-Detektion verzichtet werden.
- Einige Störungsmerkmale sind bei isolierten Vokalen besser detektierbar.

Das letztgenannte Argument wurde an der Einzelfallstudie eines Patienten mit spasmodischer Dysphonie überprüft. Im normalen Redefluß war bei diesem Patienten lediglich eine schwere Stimmstörung feststellbar, nicht aber näher spezifizierbar. Erst bei isoliert gesprochenen Vokalen fiel sofort ein starker Stimmtremor (siehe ergänzend Seite 2.2.1) als Störungsursache auf. Abbildung 4.37 zeigt den mit MODIAS bestimmten Pitch-Verlauf mit seinen auffällig periodischen Schwankungen infolge des Stimmtremors. Bei diesem Störungsbild erwies sich der MODIAS-spezifische Parameter der PITCH-VARIABILITÄT als besonders sensitiv, der um den Faktor 10 über den entsprechenden Werten sprechgesunder Probanden lag.
Zur Klärung der Frage, ob die aus isoliert gesprochenen Vokalen gewonnenen Stimmparameter auch valide sind, wurde auf ein umfangreiches Datenmaterial der Entwicklungsgruppe klinische Neuropsychologie des Städtischen Krankenhauses

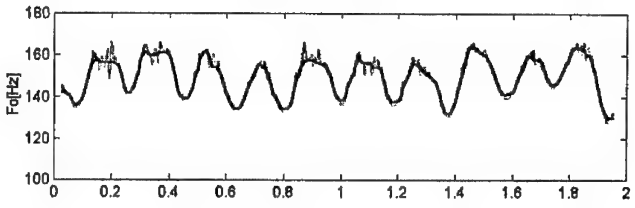

Abbildung 4.37.: Pitch-Verlauf bei einem Patienten mit schwerem Stimmtremor. Diese Störung tritt bei isolierten Vokalen wesentlich besser hervor als bei eingebetteten Vokalen im natürlichen Redefluß.

München-Bogenhausen zurückgegriffen. Eine frühere Pilotstudie im Rahmen eines Forschungsprojekts beinhaltete unter anderem die akustisch-phonetische Untersuchung von 110 Patienten mit neurogenen Stimmstörungen und 30 sprechgesunden Probanden.

Abbildung 4.38 zeigt, wie die Messungen des mittleren Pitch bei isolierten Vokalen mit den Messungen bei Vokalen zusammenhängen, die in einen Trägersatz eingebettet sind.

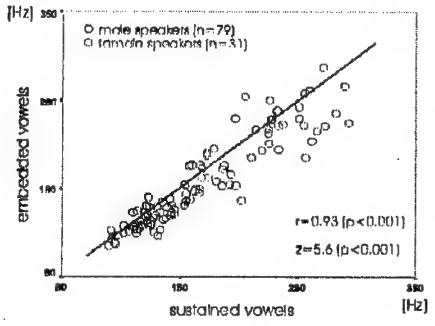

Abbildung 4.38.: Zusammenhang zwischen Messungen des mittleren Pitch an isolierten Vokalen (Abszisse) und Messungen an Vokalen, die in einen Trägersatz eingebettet sind (Ordinate). Die Korrelation ist sehr hoch, die Pitch-Messungen liegen bei isolierten Vokalen insgesamt etwas höher.

Die Korrelation ist ausgesprochen hoch. Pitch-Meßwerte für isoliert gesprochene Vokale zeigen insgesamt zwar geringfügig höhere Werte, sie schätzen jedoch hinreichend gut die korrespondierenden Meßwerte für den natürlichen Redefluß.

Abbildung 4.39 zeigt, wie gut die perzeptiven Ratings von Hörern durch solche Messungen des mittleren Pitch vorhergesagt werden können. Jeder Patient wurde von 5 Hörern bewertet, ob seine Sprechstimmlage als zu hoch (auf einer Skala von 0 bis +6) oder zu niedrig (auf einer Skala von 0 bis -6) empfunden wird. Die Pitch-Meßwerte wurden aus isolierten Vokalen abgeleitet (Abszisse), die perzeptiven Ratings hingegen aus dem natürlichen Redefluß des Patienten (Ordinate).

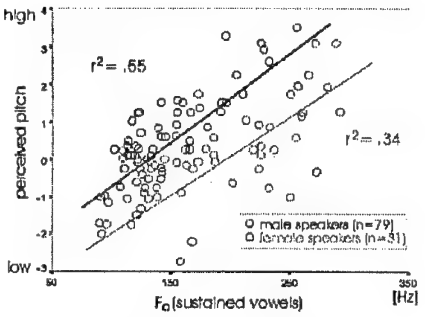

Abbildung 4.39.: Zusammenhang zwischen Messungen des mittleren Pitch an isolierten Vokalen (Abszisse) und perzeptiven Hörer-Ratings anhand des natürlichen Redeflusses (Ordinate), getrennt nach weiblichen und männlichen Patienten.

Die Korrelation zwischen akustischer Messung und perzeptivem Rating ist relativ hoch (55 bzw. 34 Prozent), gemessen an den üblicherweise zu erzielenden Korrelationen derartiger Experimente. Sie ist nur unwesentlich niedriger als bei der vergleichbaren Prädiktion aus den Meßwerten der natürlichen Sprache. Pitch-Meßwerte aus isolierten Vokalen schätzen daher hinreichend gut perzeptive Hörerurteile bezüglich der wahrgenommenen Sprechstimmlage.

4.4.2. Validität der Rauhigkeits- und Behauchtheitsprädiktion

Ob die avisierte Prädiktion der *Rauhigkeit* durch den von MODIAS bereitgestellten Parameter *JITTER* gem. Definition in 4.3.2.1 auf Seite 93, und die Prädiktion der *Behauchtheit* durch den Parameter *CPPS* gem. Definition in 4.3.2.3 auf Seite 104 mit hinreichender Sicherheit gelingt und damit von statistischer Validität im psychometrischen Sinne ausgegangen werden kann, muß im Experiment überprüft werden.
Prinzipiell ist eine solche Prüfung durch Expertenurteile unter Zuhilfenahme von Schätzskalen möglich, die aber zum einen eine hinreichend große Anzahl pathologischer Stimmbilder erfordert und zum anderen einer zeitintensiven auditiven Befundung bedarf.
Eine alternative und übliche Methode zur Prüfung der Prädiktionsleistung bietet sich durch synthetisches Stimmaterial, da hier die Stimmstörung modelliert und damit im Grad ihrer Ausprägung frei gewählt werden kann. Diese Methode wurde für MODIAS zur Prüfung der Validität herangezogen.

4.4.2.1. Erzeugung synthetischer Stimmen

Die Modellierung des menschlichen Vokaltraktes wurde bereits bei den Verfahren zur Prüfung des Artikulationsraumes hinreichend beschrieben. Hierbei wurde das aufgezeichnete Sprachsignal einer LPC-Analyse unterzogen, um die Modellparameter des Vokaltraktfilters bzw. des entsprechenden Digitalen Filters zu gewinnen.
Das Sprachsignal war bei dieser Modellierung genau dann exakt reproduzierbar, wenn das Prädiktionsfehlersignal als Anregungssignal der Stimmbänder verwendet wurde. Das Stimmband-Anregungssignal enthält aber trotz seiner vermeintlich rauschartigen Charakteristik all diejenigen sprecherspezifischen Merkmale, die die Informationen über den typischen Klang einer Stimme und ggf. auch ihrer Pathologie enthalten. Sobald das Prädiktionsfehlersignal durch ein analytisches Signal ersetzt wird, wird das resultierende Sprachsignal daher perzeptiv als künstlich wahrgenommen mit einem typisch maschinenartigen Klangbild.
Im einfachsten Falle erfolgt diese Modellierung durch einen Diracpuls mit einer dem mittleren Pitch entsprechenden Frequenz. Ein wesentlich besserer Höreindruck entsteht jedoch z.B. durch Simulation des Stimmband-Grundimpulses mit dem häufig in der Sprachverarbeitung genutzten Anregungsmodell nach ROSENBERG gem. Abbildung 4.40.

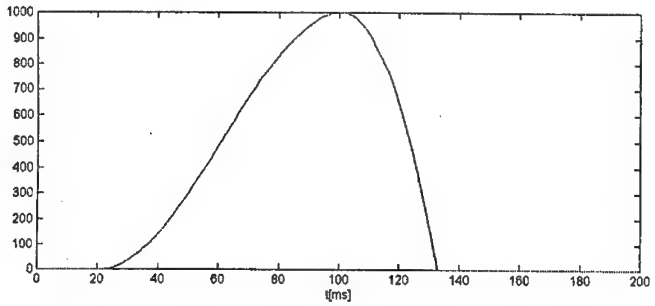

Abbildung 4.40.: Simuliertes Anregungssignal bei Verwendung des ROSENBERG-Modells. Das nach inverser Filterung eines menschlichen Sprachsignales verbleibende Anregungssignal wird entfernt und durch dieses synthetische Anregungssignal ersetzt.

Der Stimmband-Grundimpuls $G(t)$ des Rosenberg-Modells ist wie folgt definiert:

$$G(t) = \begin{cases} 1000 \cdot \left[3(\frac{t}{T_p})^2 - 2(\frac{t}{T_p})^3\right] & : \; 0 \leq t \leq T_p \\ 1000 \cdot \left[1 - (\frac{t-T_p}{T_n})^2\right] & : \; T_p < t \leq T_p + T_n \\ 0 & : \; T_p + T_n < t < T_{ges} \end{cases} \tag{4.40}$$

Die Konstanten T_p und T_n definieren dabei die relativen zeitlichen Anteile der steigenden bzw. fallenden Flanke eines Grundimpulses, bezogen auf seine Gesamtdauer T_{ges} mit der hier gewählten Dimensionierung:

$$T_p = 0.40 \cdot T_{ges} \qquad T_n = 0.16 \cdot T_{ges} \tag{4.41}$$

Die Frequenz des Stimmband-Pulses (simulierter Pitch) ergibt sich zu $F = 1/T_{ges}$.

4.4.2.2. Simulation der perzeptiven Rauhigkeit

Die Simulation einer perzeptiv wahrnehmbaren Rauhigkeit erfolgt hier durch eine Frequenzmodulation des Rosenberg-Stimmband-Pulses. Die Korrelation der simulierten Stimmstörungen mit perzeptiven Rauhigkeits-Ratings geschulter Hörer wurde im Rahmen einer Untersuchung an der Entwicklungsgruppe klinische Neuropsychologie zuvor

geprüft und bestätigt.
Die Frequenzmodulation erfolgt hier mit einer Rauschquelle, die eine innerhalb $[-R_{max} \ldots R_{max}]$ gleichverteilte Rauschamplitude R besitzt. Als indirektes Maß für den Modulationsgrad wurde ein Parameter RJ gem. einer Definition in [Sch95]) verwendet. Er ist für periodische Signale bei N zeitlich aufeinanderfolgenden Einzelschwingungen wie folgt definiert:

$$RJ = \sum_{i=2}^{N} \frac{|x_i - x_{i-1}|}{(N-1) \cdot \bar{x}} \tag{4.42}$$

Dabei ist x_i die Periodendauer der i-ten Schwingung.
Das resultierende Oszillogramm für $RJ = 0.05$ (Beispiel) ist in Abbildung 4.41 dargestellt.

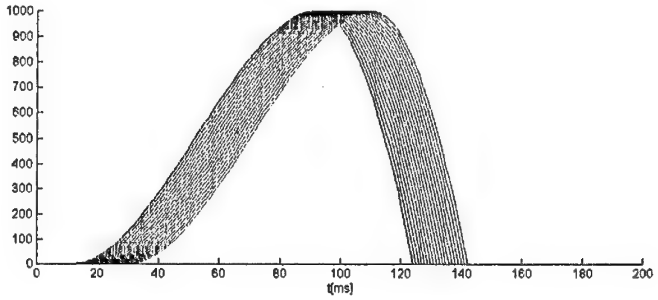

Abbildung 4.41.: Oszillogramm des simulierten Anregungssignales bei Frequenzmodulation des Rosenberg-Stimmband-Pulses (hier: $RJ = 0.05$).

4.4.2.3. Simulation der perzeptiven Behauchtheit

Die Simulation einer perzeptiv wahrnehmbaren Behauchtheit erfolgt durch Beaufschlagung des Anregungssignales mit additivem, gleichverteiltem Rauschen. Die Korrelation der simulierten Stimmstörungen mit perzeptiven Behauchtheits-Ratings geschulter Hörer wurde im Rahmen einer Untersuchung an der Entwicklungsgruppe klinische Neuropsychologie zuvor geprüft und bestätigt.

Die Nutz-Signalleistung S_{Nutz} des ungestörten (synthetischen) Anregungssignales ergibt sich durch Integration des Rosenberg-Grundimpulses $G(t)$ über eine Signalperiode T_{ges}:

$$S_{Nutz} = S_{Ros} = \frac{1}{T_{ges}} \cdot \int_{t=0}^{T_{ges}} G(t)^2 dt \tag{4.43}$$

Die Stör-Signalleistung $S_{Stör}$ einer Rauschquelle, die eine innerhalb $[-R_{max} \ldots R_{max}]$ gleichverteilte Rauschamplitude R besitzt, ergibt sich nach dem bekannten Zusammenhang zu:

$$S_{Stör} = \frac{R^2}{3} \tag{4.44}$$

Die Signalleistung des Rosenberg-Pulses gem. Gleichung 4.43 ist konstant, damit stellt sich in Abhängigkeit der frei wählbaren Rauschamplitude R ein (logarithmiertes) Signal-Störleistungsverhältnis S/N ein von:

$$S/N = 10 \cdot \log\left(\frac{S_{Nutz}}{S_{Stör}}\right) = 10 \cdot \log\left(\frac{3 \cdot S_{Ros}}{R^2}\right) \tag{4.45}$$

Die Rauschamplitude R wurde derart gewählt, daß sich ein jeweils definiertes Signal-Störleistungsverhältnis S/N in 10-dB-Schritten von $-60dB \ldots 20dB$ einstellt.
Das resultierende Oszillogramm für $S/N = 30dB$ (Beispiel) ist in Abbildung 4.42 dargestellt.

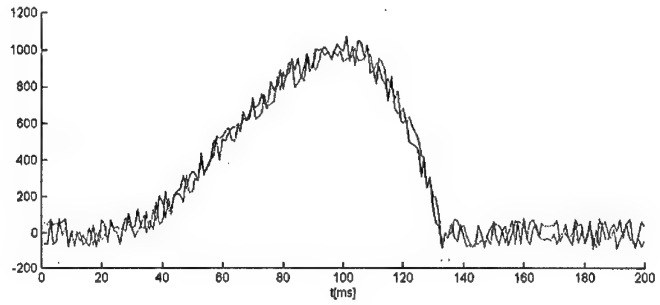

Abbildung 4.42.: Oszillogramm des simulierten Anregungssignales bei Beaufschlagung mit additivem, gleichverteiltem Rauschen (hier: $S/N = 30dB$).

4.4.2.4. Ergebnis: Korrelation mit den abgeleiteten akustischen Parametern

Rauhigkeit: Das synthetische Sprachmaterial wurde zur Prüfung des MODIAS-Rauhigkeitsmaßes *JITTER* mit schrittweise ansteigendem RJ (gem. Gleichung 4.42) beaufschlagt. Untersucht wurde hier die korrekte Reihung der Stimmproben gemäß ihrer simulierten Rauhigkeit RJ anhand der Ergebnisse der MODIAS-Stimmanalyse (*JITTER*).
Die Ergebnisse sind in Abbildung 4.43 dargestellt.

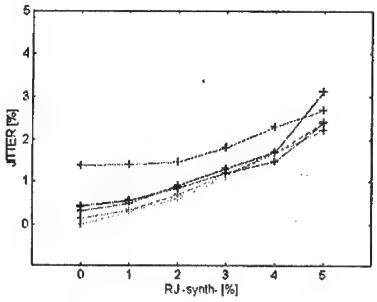

Abbildung 4.43.: Zusammenhang zwischen dem MODIAS-Rauhigkeitsmaß *JITTER* (Ordinate) und der simulierten Frequenzmodulation mit steigendem RJ (nach Gleichung 4.42). Die Kurvenschar ergibt sich durch zusätzliche Variation der simulierten Rauschstörung bei verschiedengradigem S/N (nach Gleichung 4.45).

Das MODIAS-Maß *JITTER* korreliert mit der simulierten Rauhigkeit und liefert eine korrekte Reihung der Stimmproben entsprechend dem eingestellten Störungsgrad.

Behauchtheit: Das synthetische Sprachmaterial wurde zur Prüfung des MODIAS-Behauchtheitsmaßes *CPPS* mit schrittweise ansteigendem S/N (gem. Gleichung 4.45) beaufschlagt. Untersucht wurde hier die korrekte Reihung der Stimmproben gemäß ihrer simulierten Behauchtheit S/N anhand der Ergebnisse der MODIAS-Stimmanalyse (*CPPS*). Die Ergebnisse sind in Abbildung 4.44 dargestellt.

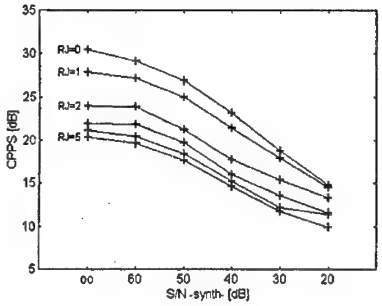

Abbildung 4.44.: Zusammenhang zwischen dem MODIAS-Behauchtheitsmaß *CPPS* (Ordinate) und der simulierten Rauschstörung steigender Rauschamplitude (d.h. fallendes S/N nach Gleichung 4.45). Die Kurvenschar ergibt sich durch zusätzliche Variation der simulierten Frequenzmodulation mit verschiedengradigem RJ (nach Gleichung 4.42).

Das MODIAS-Maß *CPPS* korreliert mit der simulierten Behauchtheit und liefert eine korrekte Reihung der Stimmproben entsprechend dem eingestellten Störungsgrad.

4.5. Anwendungen und Ergebnisse

Wie bereits dargestellt, verfolgt das Projekt MODIAS die Zielrichtung einer Standarddiagnostik neurogener Sprechstörungssyndrome. Im Rahmen der Kooperation zwischen der Klinik und Poliklinik für Mund-Kiefer-Gesichtschirurgie der Technischen Universität München, Klinikum rechts der Isar, mit der Entwicklungsgruppe klinische Neuropsychologie des Städtischen Krankenhauses München-Bogenhausen konnte aber auch ein fruchtbarer und gewinnbringender Einsatz des MODIAS-Systems bei der apparativ phonetischen Untersuchung an Patienten mit Lippen-Kiefer-Gaumenspalten, also einer völlig anderen Zielgruppe gezeigt werden. Die Untersuchung dieser insgesamt 154 Patienten wurde ausführlich in [Bre98] publiziert; das Ergebnis bzw. die Schlußfolgerung der Untersuchung lautete wie folgt:

- Während bisherige Vermutungen von einer erhöhten Vulnerabilität von LKG-

Patienten für Stimmstörungen ausgegangen sind, war die in unserer Untersuchung beobachtete Prävalenz von Stimmstörungen gegenüber der Normalbevölkerung nur geringgradig erhöht.

- Die apparative akustische Analyse kann in der klinischen Einzelfalldiagnostik eine sinnvolle und praktikable Ergänzung der perzeptiven Befundung sein.

Eine hingegen idealtypische Anwendung des Moduls STIMME UND VOKALARTIKULATION bestand in einer 1999 durchgeführten Untersuchung der Entwicklungsgruppe klinische Neuropsychologie des Städtischen Krankenhauses München-Bogenhausen. Im Rahmen dieser Studie wurden die Stimmparameter von 10 Patienten mit neurogenen Stimmstörungen verglichen mit den entsprechenden Parametern von 12 Normalsprechern. Allgemein konnte mit dieser Studie gezeigt werden, daß sich bei jedem Patienten mit perzeptiv wahrnehmbaren Stimmstörungen mindestens einer der MODIAS-Stimmparameter als sensitiv herausstellte. Damit war in allen Fällen die Stimmstörung ausschließlich auf Basis der akustischen Analysen diagnostizierbar.
Abbildung 4.45 zeigt die Ergebnisse der Pitch-Verlaufsanalyse.

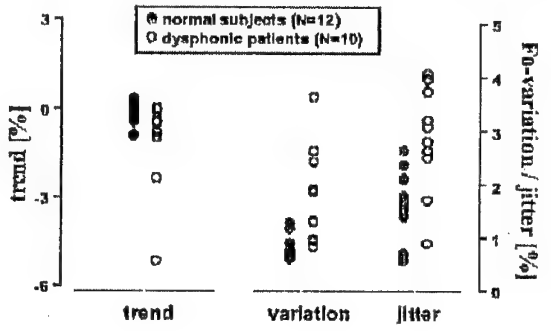

Abbildung 4.45.: Ergebnisse der Pitch-Verlaufsanalyse für 10 Patienten mit neurogenen Stimmstörungen und 12 Normalsprecher.

Der Parameter PITCH-TREND zeigte bei zwei Patienten signifikant negative Werte, entsprechend einem zeitlich stark abfallenden Pitchverlauf von bis zu 5 Prozent pro Sekunde. Beide Patienten litten an einer Schwäche der Kehlkopfmuskulatur. Bei 6 von 10

Patienten traten überhöhte Werte für die Parameter PITCH-VARIABILITÄT und JITTER auf. Abbildung 4.46 zeigt die Ergebnisse der CPPS-Verlaufsanalyse.

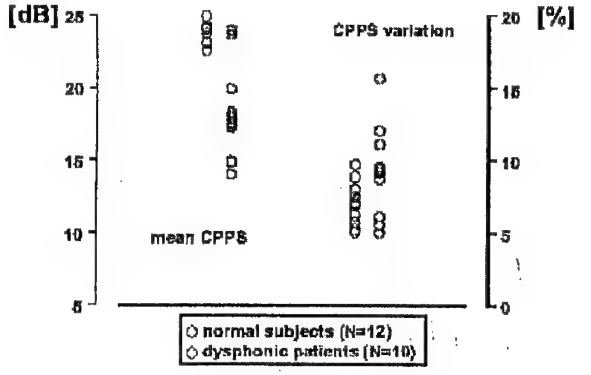

Abbildung 4.46.: Ergebnisse der CPPS-Verlaufsanalyse für 10 Patienten mit neurogenen Stimmstörungen und 12 Normalsprecher.

Der Parameter MITTLERE CPPS zeigte bei fast allen Patienten reduzierte Werte, offensichtlich verursacht durch einen unvollständigen Glottis-Verschluß und den damit erhöhten Behauchungsgrad. Bei 3 von 10 Patienten war der Parameter CPPS-VARIABILITÄT erhöht, verursacht durch intermittierende Einbrüche der Stimmqualität während der anhaltenden Vokalisation.
Die Ergebnisse der Studie belegen, daß die Sensitivität des MODIAS-Moduls für die Detektion neurogener Stimmstörungen ausreicht. Die Spezifität des Systems zur Klassifizierung der erkannten Stimmstörungen muß noch in nachfolgenden Studien gezeigt werden (Sämtliche Ergebnisse sind in [Merk99b] publiziert).

5. Realisierung des Moduls SATZPRODUKTION

Das diagnostische und technische Konzept zu dem Modul SATZPRODUKTION wurde bereits im Kapitel 3 dargestellt. Nach einer kurzen Erläuterung der relevanten diagnostischen Fragestellungen widmet sich dieses Kapitel nun der konkreten Realisierung innerhalb des MODIAS-Systems. Der interdisziplinäre Charakter der Aufgabenstellung läßt hierbei eine Zweiteilung sinnvoll erscheinen in die

- klinisch/diagnostisch relevanten Anteile, wie z.B. Untersuchungsprotokoll, gewähltes Untersuchungsmaterial, Ablauf einer Untersuchungssitzung und Auswertung bzw. Darstellung der Ergebnisse, sowie in

- technisch relevante Anteile, wie z.B. angewandte bzw. entwickelte technische Methoden, Verfahren und Algorithmen, die zu der ingenieurwissenschaftlichen Lösung der medizinischen Problemstellung geführt haben.

Der letzte Abschnitt zeigt eine klinische Anwendung des Moduls SATZPRODUKTION im Rahmen einer Langzeit-Therapiekontrolle an 3 Patienten mit schweren neurogenen Spechstörungen. Die Untersuchung wurde in der Entwicklungsgruppe klinische Neuropsychologie der Abteilung für Neuropsychologie am Städtischen Krankenhaus München-Bogenhausen durchgeführt.

5.1. Diagnostische Fragestellungen

Ein verändertes Sprechtempo gehört zu den häufigsten Symptomen neurogener Sprechstörungen. Sowohl Abweichungen nach oben als auch nach unten werden beobachtet, ein verlangsamtes Sprechtempo tritt jedoch in praxi bei weitem häufiger auf ([Ack92];[Zie88];[Zie93c]).
Von besonderer diagnostischer Bedeutung ist die Frage der Verlangsamung bei Patienten mit aphasischen und mit sprechapraktischen Störungen. In diesen Fällen ist die Reduktion des Sprechtempos nicht durch eine elementar-motorische Störung bedingt, sondern Ausdruck einer Beeinträchtigung höherer Prozesse der Sprachproduktion, die darüber hinaus auch durch segmentale Fehler charakterisiert ist.
Unter den segmentalen Fehlern sind die verschiedenen Typen phonematischer Paraphasien von Bedeutung, suprasegmental werden Fehlversuche, Repairs, Iterationen, Pausen, Akzentfehler etc. beschrieben. Während die Analyse segmentaler Fehler nach auditiven Kriterien ein klinisch etabliertes Verfahren darstellt, gibt es für die Quantifizierung der Zeitaspekte gestörter Sprachproduktion noch kein klinisch anwendbares Verfahren ([Merk97a],[Merk97b]).
Hier wird eine PC-gesteuerte Methode zur akustischen Analyse des Zeitmusters der Satzproduktion bei Patienten mit neurogenen Sprechstörungen beschrieben. Das Verfahren erlaubt gleichzeitig eine auditive Analyse nach phonetischen und phonematischen Fehlern. Es eignet sich zur Differenzierung unterschiedlicher Störungsmuster und zur Analyse der Faktoren, die das Auftreten von Unflüssigkeiten beeinflussen.

5.2. Aufbau und Gestaltung

5.2.1. Untersuchungsprotokoll und -material

Der Patient hat in diesem Modul die Aufgabe, 24 ausgewählte Testsätze nachzusprechen. Die Untersuchung kann in zwei unterschiedlichen Modi durchgeführt werden:

- Modus 1: Schwerpunkt *Diagnostik.* Die 24 Sätze bestehen aus einer konstanten Trägerphrase (*Ute kann die ... bekommen*) mit je einem eingebetteten Zielwort.

- Modus 2: Schwerpunkt *Therapiekontrolle*. Hier werden zusätzlich die Trägersätze variiert bei näherungsweise gleichem Satzrhythmus und identischer Zielwortmenge (z.B. *Opa darf die ... benutzen, Ida hat die ... genommen*).

Die Zielwörter haben unterschiedliche phonologische Strukturen und sind sorgfältig phonetisch balanciert ([Bre97]). Sie bestehen zu gleichen Teilen aus

- zweisilbigen Wörtern mit einfacher Konsonant-Vokal-Struktur, z.B. *Tüte*,
- zweisilbigen Wörtern mit Konsonant-Cluster, z.B. *Strümpfe*,
- dreisilbigen Wörtern, z.B. *Trompete*.

Somit entstehen insgesamt jeweils 8 Sätze pro Zielworttyp bzw. Zielwortkomplexität.
Der erste Modus ist speziell für die (einmalige oder zeitlich deutlich abgesetzte) Diagnostik konzipiert mit besonderem Augenmerk auf der Beobachtung von Zögerphänomenen vor dem betreffenden Zielwort. Die Trägerphrase ist hier über alle 24 Testsätze konstant gehalten, die einzige Variation findet an der Stelle des Zielwortes statt. Sprechapraktisch gestörte Patienten erlernen relativ schnell die Produktion des stets gleichen Trägersatzes; Probleme stellen sich erst am Beginn der Variation, also am eingefügten und meist komplexeren Zielwort ein. An dieser Stelle treten Zögerphänomene, Stottern oder völlige Abbrüche der Satzproduktion auf. Der zweite Modus ist leicht abgewandelt und daher besser für die Baseline-Phase einer Therapiekontrolle geeignet (Wiederholte Untersuchung an aufeinanderfolgenden Tagen unmittelbar vor Beginn der Therapie zur Ermittlung von Meßwertstreuungen). Hier werden die Trägersätze variiert, um Lerneffekte zu vermeiden.
Die am Städtischen Krankenhaus München-Bogenhausen durchgeführten Beobachtungen von Therapieverläufen allein mit dem Untersuchungsmodus 1 hatten gezeigt, daß sich die Patienten in der Baseline-Phase der Therapiekontrolle zunächst deutlich in ihrer Satzproduktion verbesserten, was sich sowohl in der akustischen als auch in der auditiven Analyse manifestierte. Anschließend erfolgte für jeden Patienten eine intensive, mehrwöchige logopädische Therapie, deren Erfolg abschließend wieder durch mehrere Untersuchungsgänge geprüft werden sollte. Dabei stellte sich heraus, daß die jetzt durchgeführte akustische Analyse keine Verbesserung gegenüber der letzten Untersuchung (also vor der Therapie)

zeigte, sondern vielmehr eine geringfügige Verschlechterung. Diese akustisch basierte Beobachtung deckte sich jedoch wie erwartet nicht mit dem perzeptiv gewonnenen Bild. Bei der Konzeption des Untersuchungsprotokolles war man zunächst davon ausgegangen, daß sich eine Sättigung des Lerneffektes schon nach wenigen Sätzen innerhalb nur eines Untersuchungsganges einstellen würde. Die Praxis zeigte aber sehr schnell, daß sich dieser Lerneffekt auch über mehrere Untersuchungsgänge hinaus beobachten ließ, was bei der (einmaligen) Diagnostik zwar unkritisch, jedoch bei einer engmaschigen Verlaufskontrolle nicht mehr akzeptabel ist. Da bei der Kontrolle therapeutischer Effekte die Information über die Veränderung von einem Untersuchungsgang zum nächsten von Interesse ist, dürfen die verwendeten diagnostischen Verfahren nicht durch derartige Lerneffekte verfälscht werden. Die durchgeführten Folgeversuche konnten unmittelbar zeigen, daß die beschriebene Diskrepanz zwischen akustischen Daten und perzeptivem Korrelat bei neuerlichen Therapieüberwachungen immer dann nicht mehr zu beobachten war, wenn auch die Trägerphrasen variiert wurden.
Das Untersuchungsprotokoll des Moduls SATZPRODUKTION sieht grundsätzlich nur 1 Durchgang vor und damit auch die Entscheidung für einen der beiden Modi. Insgesamt entstehen daher pro Untersuchung 24 Satzaufzeichnungen bzw. jeweils 8 Satzaufzeichnungen pro Zielwortkomplexität.

5.2.2. Untersuchungsablauf und Bedienung

Nach dem Aufruf des MODIAS-Systems erscheint die Startoberfläche gem. Abbildung 4.1 auf Seite 47 mit dem entsprechenden Startbutton für das Modul SATZPRODUKTION.
Nach Betätigung des Buttons START! gelangt man unmittelbar zur Organisationsoberfläche gem. Abbildung 4.2 auf Seite 48 mit der Aufforderung zur Eingabe sämtlicher relevanter Daten zum Patienten und zur aktuellen Untersuchungssitzung.
Der Untersucher wird jetzt zur Steuerungsoberfläche gem. Abbildung 5.1 weitergeleitet, die eine Festlegung bezüglich des weiteren Fortgangs der Untersuchung fordert bzw. ermöglicht.
Im oberen Teil der Steuerungsoberfläche wird der bereits dargestellte Untersuchungsmodus festgelegt, das ist entweder

- Modus 1 mit festen Trägerphrasen bei variiertem Zielwort, daher optimiert für die

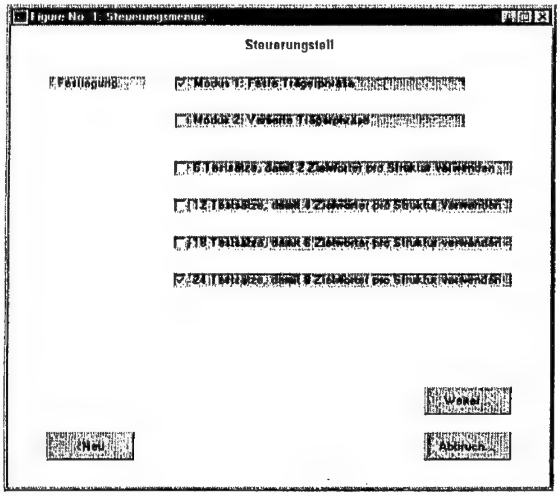

Abbildung 5.1.: Steuerungsoberfläche des Moduls SATZPRODUKTION zur Festlegung des weiteren Untersuchungsablaufs. An dieser Stelle wird vom Untersucher angegeben, welcher Untersuchungsmodus (feste oder variierte Trägerphrase) bzw. wieviele Testsätze innerhalb des gewählten Modus genutzt werden sollen.

Beobachtung von Zögerphänomenen beim Übergang von der bekannten Trägerphrase zum unbekannten und artikulatorisch komplexer zu realisierenden Zielwort (Schwerpunkt: Diagnostik),

- Modus 2 mit zusätzlich variierenden Trägerphrasen, daher optimiert für die Beobachtung von längerfristigen Veränderungen bzw. mehreren Untersuchungen am gleichen Patienten (Schwerpunkt: Engmaschige Verlaufskontrolle).

Im unteren Teil der Steuerungsoberfläche kann der Untersucher auswählen, ob das

- komplette Standardprotokoll mit 24 Testsätzen und damit je 8 Testsätzen pro Zielwortkomplexität oder aber das
- verkürzte Untersuchungsprotokoll mit 18, 12 oder nur 6 Testsätzen

genutzt werden soll.

Durch Betätigung des Buttons WEITER... gelangt man in die erste der bis zu 4 aufeinanderfolgenden Aufzeichnungsoberflächen gem. Abbildung 5.2. Diese Verteilung ergibt sich in Abhängigkeit von der Länge des Untersuchungsprotokolls. Die Darstellung der Testsätze erfolgt bezüglich der enthaltenen Zielwortkomplexität randomisiert.

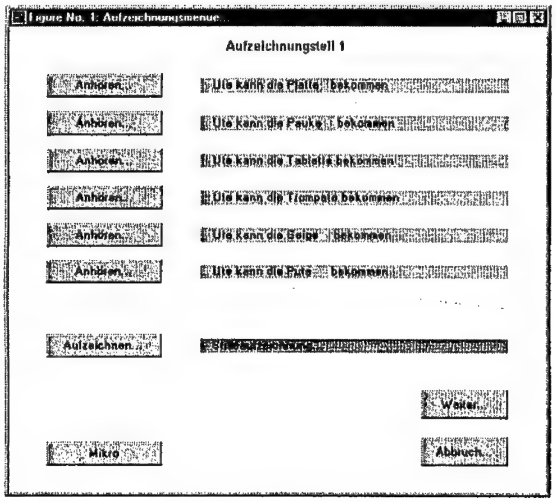

Abbildung 5.2.: Eine der bis zu 4 aufeinanderfolgenden Aufzeichnungsoberflächen des Moduls SATZPRODUKTION. Die Testsätze sind bezüglich der enthaltenen Zielwortkomplexität randomisiert. Die Aufzeichnung wird jeweils durch den Button AUFZEICHNEN... gestartet. Liegt bereits eine Aufzeichnung vor, ist nur die Wiedergabe über den vom System ersetzten Button ANHÖREN... möglich.

Die Aufzeichnung jedes Satzes wird durch Betätigung des entsprechenden Buttons AUFNEHMEN... gestartet.

Bei der Aufzeichnung wird jeder Testsatz vom Patienten in normalem Tempo möglichst natürlich nachgesprochen. Soll die Aufzeichnung gespeichert werden, genügt das Schließen des Soundrecorders (siehe Abbildung 4.5 auf Seite 51) und die Speicherung wie vorgeschlagen. Der Dateiname ist durch MODIAS bereits voreingestellt, ebenso die Sy-

stemeinstellungen bezüglich Samplingrate (22050 Hz) und Auflösung (16Bit).
Die im Untersuchungsprotokoll vorgesehene Stilleaufzeichnung sieht eine Aufzeichnung ohne Aufsprache bzw. ohne jegliche Art von (nicht systemimmanenten) Störgeräuschen mit einer Dauer von ca. 5 sec vor. Eine automatisch durchgeführte statistische Analyse des systemspezifischen Rauschens des Aufzeichnungssystems Soundkarte/Rechner ermöglicht eine wesentlich exaktere Festlegung der Silbengrenzen bzw. von Signalabschnitten mit artikulatorischer Aktivität.
Wie bereits dargestellt, besteht jede Aufzeichnung aus einem neun- bzw. zehnsilbigen Satz, zwischen den Silben können Sprechpausen unterschiedlicher Dauer liegen. Zur Ableitung der suprasegmentalen bzw. zeitlichen Aspekte der Satzproduktion ist es zunächst erforderlich, den aufgezeichneten Satz weitgehend automatisch in einzelne Silben zu segmentieren. Da sämtliche hierzu bekannten Verfahren bei unscharfer Artikulation an ihre Grenzen stoßen, muß neben einem gesteigerten Aufwand in das technische Verfahren dem Untersucher auch die Möglichkeit gegeben werden, in die Segmentierung steuernd eingreifen zu können. Gerade in der zeitlichen Lage der Einzelsilben und in ihrer Dauer liegt die Information, die zur späteren Beurteilung des Redeflusses herangezogen werden soll.
Hierzu wurde in dem Modul SATZPRODUKTION ein spezieller Segmentierungseditor implementiert, der einen mit der subjektiven Lautheitsempfindung des menschlichen Gehörs korrespondierenden Kurvenverlauf (Lautheitskontur) darstellt und damit eine geeignete Oberfläche für die weitere Bearbeitung durch den Untersucher zur Verfügung stellt. Die erforderlichen Berechnungen der Lautheitskonturen sowie der daraus abgeleiteten Silbengrenzen werden automatisch beim Übergang von den Aufzeichnungsoberflächen zur anschließenden Segmentierungsoberfläche gem. Abbildung 5.3 für jeden der realisierten Testsätze berechnet.
Die Segmentierungsoberfläche gestattet für jede durchgeführte Aufzeichnung (erkennbar an einer gelben Markierung am rechten Rand der jeweiligen Textfelder) über den Button SEGMENTIEREN... den Aufruf des Segmentierungseditors gem. Abbildung 5.4.
Im Segmentierungseditor werden zunächst die berechneten Lautheitskonturen für den jeweiligen Testsatzes grafisch darstellt und dabei die vom Segmentierungsalgorithmus automatisch erkannten Silben entsprechend markiert. Die Lautheitskonturen zeigen idealtypisch lokale Maxima an den Orten der Silbenkerne bzw. lokale Minima an den Orten der Silbenübergänge bzw. in den Sprechpausen.
Die rechts oben eingeblendeten Informationsfelder zeigen dem Untersucher den aktuellen

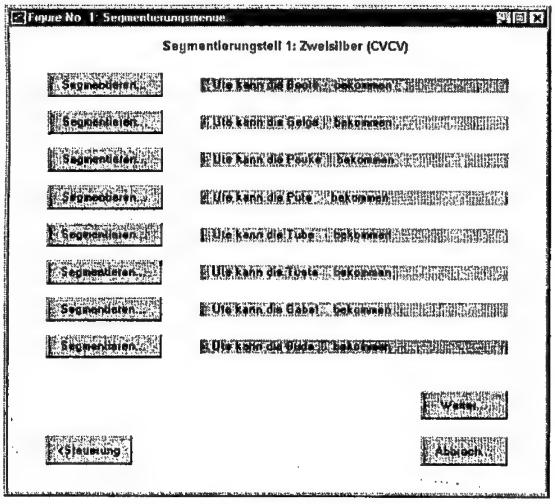

Abbildung 5.3.: Eine der 3 Segmentierungsoberflächen des Moduls SATZPRODUKTION, hier für die Komplexität zweisilbiger Zielworte mit einfacher Konsonant-Vokal-Struktur (CVCV).

Stand der Silbensegmentierung für den gerade bearbeiteten Testsatz an. Diese Felder sind dabei folgendermaßen zu interpretieren:

- ZIELSILBEN: Soll-Anzahl der (Ziel-)Silben des nachzusprechenden Testsatzes. Sie ist a priori bekannt und hängt ausschließlich von der Zielwortkomplexität ab (Wert 9 bei zweisilbigem Zielwort bzw. Wert 10 bei dreisilbigem Zielwort).

- BISHER ERMITTELT: Ist-Anzahl der Zielsilben des tatsächlich produzierten Testsatzes nach dem gegenwärtigen Kenntnisstand des Systems. Zielsilben sind für alle weiteren Betrachtungen die vom Patienten produzierten Silben, die auch tatsächlich zum Silbenumfang des Testsatzes gehören. Die Anzeige ist dynamisch und hängt sowohl von der akustisch basierten Silbensegmentierung als auch von der auditiv basierten Silbenattribuierung (Erläuterung hierzu auf Seite 125) ab.

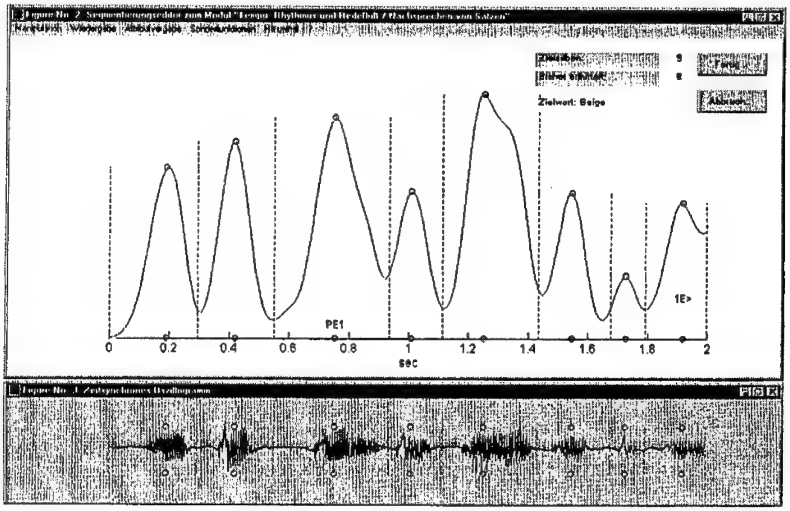

Abbildung 5.4.: Bedienoberfläche des Segmentierungseditors des Moduls SATZPRODUKTION. Hier erfolgt die manuelle Einflußnahme des Untersuchers auf die vom System vorgeschlagene Segmentierung wie auch die Eingabe auditiver Störungsmerkmale (Silbenattribute). Zusätzlich wird das zeitsynchrone Oszillogramm eingeblendet.

Müssen die vorgeschlagenen Silbengrenzen bezüglich ihrer zeitlichen Lage korrigiert werden, stehen dem Untersucher mehrere Interaktionsmöglichkeiten zur Verfügung. Unterstützend wird hierzu im unteren Teil des Bildschirms das zeitsynchrone Oszillogramm zur ergänzenden Visualisierung des Sprachsignales eingeblendet.
Das Menü MANIPULATION dient der unmittelbaren, manuellen Einflußnahme auf die bisher automatisch erfolgte Segmentierung. Es bietet die folgende Auswahl:

- ZWEI SILBEN VEREINIGEN: Diese Funktion ist immer dann sinnvoll, wenn eine Silbe z.B. aufgrund eines Lautheitseinbruches im Silbenkern vom Segmentierungsalgorithmus fälschlicherweise in zwei (Halb-)Silben geteilt wurde. Das kann vor allem bei Plosiv-Vokal-Folgen mit ausgeprägtem Verschlußlösungsgeräusch geschehen. Hierzu Silbenkern mit linker Mousetaste selektieren; damit wird die aktuelle

Silbe mit der unmittelbar nachfolgenden Silbe vereinigt.

- SILBE LÖSCHEN: Gestattet die Entfernung von Silben, z.B. bei fehlinterpretierten Artefakten. Der betreffende Silbenkern wird mit linker Mousetaste selektiert, damit wird die gesamte Silbe entfernt.
- SILBE SETZEN: Manuelle Markierung einer Silbe, die ggf. vom Algorithmus zunächst unerkannt geblieben war. Hierzu Silbenkern mit linker Mousetaste setzen; die zugehörigen Silbengrenzen werden automatisch vom System vorgeschlagen.
- LINKE bzw. RECHTE SILBENGRENZE VERSCHIEBEN: Erlaubt die manuelle Verschiebung von Silbengrenzen. Der Silbenkern wird hierfür mit der linken Mousetaste selektiert, danach wird der neue Ort der linken bzw. rechten Silbengrenze mit gleicher Mousetaste vorgegeben.

Das Menü WIEDERGABE gestattet dem Untersucher die perzeptive Kontrolle der bisherigen Silbensegmentierung. Die Auswahl erlaubt sowohl die Wiedergabe der gesamten Aufzeichnung als auch die Wiedergabe ausschließlich der segmentierten Silben innerhalb ihrer jeweiligen Grenzen. Die Wiedergabe einzelner Silben ist ebenfalls vorgesehen.
Die zusätzliche Auswahl der ZIELSILBENWIEDERGABE sieht die Beschränkung der akustischen Darbietung auf die bereits beschriebenen Zielsilben vor, also diejenigen Silben, die auch tatsächlich zum Silbenumfang des zu produzierenden Testsatzes gehören. Jede weitere Silbe trägt aufgrund ihrer Redundanz (wiederholte Silben) bzw. Irrelevanz (hinzugefügte Silben) nicht zur Informationsübertragung bei und ist in diesem Sinne überflüssig.

Das Menü ATTRIBUTVERGABE stellt dem Untersucher eine Schnittstelle zur Verfügung, bei Bedarf jede beliebige Silbe mit zusätzlicher auditiver Information zu versehen. Die Attribuierung einer segmentierten Silbe erfolgt durch Auswahl des avisierten Attributes und durch Plazierung mit der linken Mousetaste direkt am jeweiligen Silbenkern. In einem ersten Schritt werden diejenigen Silben mit Attributen versehen, die nicht zum eigentlichen Testsatz gehören. Das sind:

- IT (ITERATION): Wiederholte Silbe z.B. durch Stottern;
- AD (HINZUFÜGUNG): Addierte Silbe, Fehlversuch oder Selbstkorrektur.

In einem zweiten Bearbeitungsschritt werden bei Bedarf (Ziel-)Silben markiert, die vom Patienten gar nicht realisiert worden sind. Da in diesem Falle logischerweise keine direkte Attribuierung möglich ist, wird die unmittelbar vorhergehende Silbe mit dem Attribut

- E> (ELISION): Nachgeordnet wurden Zielsilben nicht produziert -

versehen. An dieser Stelle stehen dem Untersucher die Attribute 1E> (d.h. 1 ausgelassene Silbe folgt) bis 9E> (d.h. 9 ausgelassene Silben folgen) zur Verfügung. Das Attribut 9E> würde beispielsweise dann vergeben, wenn im Falle eines 10-silbigen Testsatzes schon nach der ersten Silbe ein völliger Abbruch der Satzproduktion stattfände.
Sinn und Zweck der bisherigen Attribuierung gem. Schritt 1 und 2 ist es, dem MODIAS-System eine eindeutige Zuordnung zwischen den detektierten Maxima der Lautheitskontur und den perzeptiv wahrnehmbaren Silben treffen zu können, also zwischen den realisierten Silben (einschließlich vorkommender Iterationen, Additionen und Elisionen) und den 9 bzw. 10 tatsächlichen Zielsilben. Diese Aufgabenstellung ist prinzipiell durch eine vollständige phonetische Transkription sämtlicher realisierter Silben lösbar.
Einfacher und effizienter erscheint hier der Ansatz, den Untersucher nicht generell zur Eingabe auditiver Information aufzufordern, sondern nur dann, wenn diagnostisch relevante Abweichungen von einer ungestörten Satzproduktion auftreten. Der Segmentierungseditor ist daher so konzipiert, daß sich die geforderte Zuordnung bei abgeschlossener und fehlerfreier Attribuierung automatisch ergeben muß, sie wird ständig algorithmisch und damit nicht sichtbar für den Untersucher überwacht.
Das bereits beschriebene Informationsfeld BISHER ERMITTELT wird bei Vergabe der Attribute IT bzw. AD jeweils dekrementiert, da die soeben markierte Silbe keine Zielsilbe darstellt und damit aus der Menge der betrachteten Zielsilbenkandidaten herausfällt. Bei Vergabe des Attributes E> findet keine sichtbare Veränderung des Informationsfeldes statt, jedoch wird die Information bezüglich der fehlenden Silben bei der Plausibilitätsprüfung am Ende des Segmentierungsvorganges entsprechend berücksichtigt (siehe Seite 127).
Im dritten und letzten Schritt werden diejenigen Silben markiert, die gemäß perzeptivem Höreindruck des Untersuchers bzw. seiner auditiven Befundung segmentale Fehler beinhalten. Möglich sind hier

- PE (PHONETISCHE ENTSTELLUNG): Silbe beinhaltet in der phonetischen Ausprägung eine diagnostisch relevante Lautveränderung. Hierbei wird präzisierend

unterschieden, ob Silbenonset (anklingende Halbsilbe) oder Rhyme (abklingende Halbsilbe) betroffen sind (Attribut PE1) oder sogar beide Halbsilben (Attribut PE2).

- PP (PHONEMATISCHE PARAPHASIE): Silbe wurde durch Ersetzung, Auslassung, Umstellung oder Hinzufügung einzelner Laute verändert. Auch hier wird unterschieden, ob Silbenonset oder Rhyme (Attribut PP1) bzw. beide betroffen sind (Attribut PP2).

Sollte es aufgrund einer verwaschenen bzw. wenig kontrastreichen Artikulation des Patienten oder auch aufgrund koartikulatorischer Effekte (hier besonders häufig bei den letzten beiden Silben *../kom/men/*) für den automatischen Silbensegmentierungsalgorithmus nicht möglich sein, zwei benachbarte Silben zu trennen, kann zusätzlich das Attribut

- VE (VERSCHMELZUNG): Markierte Silbe besteht eigentlich aus 2 Einzelsilben und konnte nicht getrennt werden -

vergeben werden. Damit wird dieses Silbenkonglomerat intern wie 2 getrennte Silben mit jeweils halber Silbendauer betrachtet. Das Informationsfeld BISHER ERMITTELT wird entsprechend inkrementiert.

Durch die gewählte Art der Silbenattribuierung und die parallele Kontrolle über den beschriebenen Silbenzähler ist es nun algorithmisch möglich, neben der gesuchten eindeutigen Zuordnung der Zielsilben auch die vom Untersucher eingegebene auditive Information in gewissen Grenzen auf ihre Plausibilität hin zu überprüfen. Aus logischer Überlegung folgt, daß nach fehlerfreier Durchführung sowohl der Segmentierung als auch der Attribuierung die beiden erläuterten Informationsfelder in einem direkten Zusammenhang stehen müssen: Der Feldinhalt BISHER ERMITTELT entspricht zu diesem Zeitpunkt dem Feldinhalt ZIELSILBEN, bzw. ist er beim Vorkommen von Elisionen um deren Summe erhöht. Der Segmentierungseditor kann softwaregesteuert nur bei Erfüllung dieses Plausibilitätskriteriums mit dem Button OK verlassen werden. Jede Abweichung wäre an dieser Stelle nur auf einen (oder mehrere) Attribuierungsfehler des Untersuchers zurückzuführen, was aufgrund der besonders hohen diagnostischen

Relevanz der auditiven Information zu folgenschweren Fehlinterpretationen führen könnte.

Das Menü SONDERFUNKTIONEN bietet dem Nutzer die Möglichkeit, den Algorithmus zur Berechnung der Lautheitskontur an die jeweilige Sprechgeschwindigkeit des Patienten anzupassen. Werden bei Aufruf des Segmentierungseditors zunächst weniger als 9 bzw. 10 Silben detektiert und liegt subjektiv ein erhöhtes Sprechtempo vor, kann hier die Auswahl ANPASSUNG AN HOHE SPRECHGESCHWINDIGKEIT erfolgversprechend sein. Werden umgekehrt mehr als 9 bzw. 10 Silben erkannt und liegt subjektiv ein verlangsamtes Sprechtempo vor, sollte die Anpassung der Sprechgeschwindigkeit entsprechend nach unten erfolgen.

Die Auswahl MANUELLE THRESHOLD-FESTLEGUNG gestattet dem Untersucher die spezifische Vorgabe einer Mindestlautheit für die Detektion von Silben direkt in der Grafik über die linke Mousetaste.

Das Menü HILFSMITTEL ermöglicht eine benutzerdefinierte Zoomfunktion bezüglich der Zeitachse bei fester Skalierung der Lautheitsachse. Die Festlegung des darzustellenden Bereiches erfolgt durch Aufziehen eines Zoomfensters bei gedrückter linker Mousetaste. Die drei weiteren Auswahlmöglichkeiten dieses Menüs steuern das Ein- bzw. Ausblenden des Oszillogrammes sowie die Anpassung der Oszillogramm-Darstellung auf bereits erfolgte Manipulationen (linke bzw. rechte Silbengrenze verschoben) oder geänderte Skalierungen (Zeitachse gezoomt). Das letzte Menü ? bietet eine Hilfefunktion mit ausführlicher Beschreibung der Bedienung des Segmentierungseditors.

Nach Verlassen des Segmentierungseditors mit dem Button OK werden die jetzt überflüssigen Signalabschnitte vor der ersten Silbe und nach der letzten Silbe entfernt. Sind alle Aufzeichnungen innerhalb der Segmentierungsoberfläche bearbeitet, erfolgt der Übergang zur Auswertung.

5.2.3. Auswertung, Ergebnisdarstellung und Datenexport

Die Auswertungsoberfläche gem. Abbildung 5.5 gestattet die getrennte Auswertung nach den 3 verschiedenen Zielwortkomplexitäten, hier bezeichnet als

- Zweisilbler,
- Komplexe Zweisilbler,
- Dreisilbler,

sowie die übergreifende Auswertung über alle 24 Testsätze.

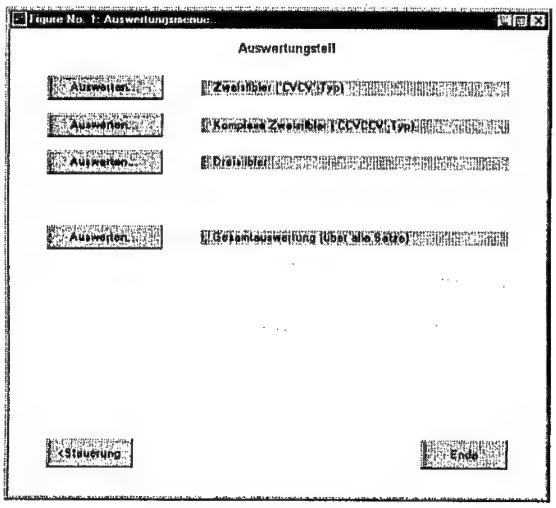

Abbildung 5.5.: Auswertungsoberfläche des Moduls SATZPRODUKTION. Von hier aus werden die Einzelauswertungen nach Zielwortkomplexität bzw. die übergreifende Auswertung über alle Testsätze initiiert.

5.2.3.1. Auswertung getrennt nach Zielwortkomplexität

Das erste Auswertungsblatt gem. Abbildung 5.6 widmet sich primär der grafischen Visualisierung von Tempo, Rhythmus und Redefluß, wobei die akustisch gewonnene Information mit der auditiven Information des Untersuchers geeignet verschränkt werden soll. Hierzu wurde für beide Grafiken eine Darstellung gewählt, die auf der Abszisse ein festes, äquidistantes Raster mit 9 bzw. 10 Zielsilben enthält.

Im oberen Teil des Auswertungsblattes sind für jeden der 8 Testsätze der gewählten Zielwortkomplexität die ermittelten Silbendauern in der Dimension ms über der entsprechenden Silbe aufgetragen. Die Silbendauer ist die zeitliche Differenz zwischen jeweils rechter und linker Silbengrenze. Der resultierende Verlauf jedes Testsatzes ist hier in einer eigenen Farbe dargestellt gemäß der unten rechts dargestellten Legende.

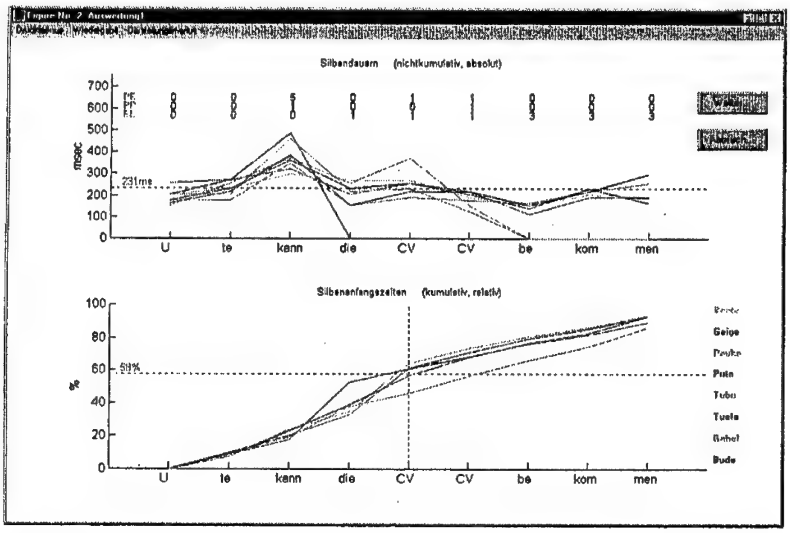

Abbildung 5.6.: Auswertung zum Modul SATZPRODUKTION, getrennt nach Zielwortkomplexität, Blatt 1. Die gewählte Art der Darstellung visualisiert grafisch den Sprechfluß und deckt Zögerzeiten (häufig vor dem Zielwort) auf.

Das hier gezeigte Beispiel gem. Abbildung 5.6 entstammt einer Untersuchung an einem Patienten mit Sprechapraxie-Syndrom. Bis zum Zielwort ist die Satzproduktion überwiegend unauffällig. Der Testsatz mit Zielwort *Pute* wurde hier schon vor Beginn des Zielwortes völlig abgebrochen, erkennbar an den folgenden Silbendauern mit dem Wert Null.
Die eingeblendete gestrichelte Linie zeigt den Durchschnittswert der Silbendauern an, berechnet über alle realisierten Silben (Grand Average), hier 231ms.

Die am oberen Rand dargestellte Tabelle gibt an, bei welcher Silbe und dort mit welcher absoluten Häufigkeit phonetische Entstellungen (PE), Phonematische Paraphasien (PP) und Elisionen (EL) innerhalb der realisierten Sätze festgestellt wurden. Damit eröffnet sich dem Untersucher die Möglichkeit, quantifizierbare Aussagen bezüglich der Ursachen unflüssiger Satzproduktion zu treffen.
Im unteren Teil des Auswertungsblattes sind die Silbenstartzeitpunkte in ihrer relativen zeitlichen Lage bezogen auf die Gesamtdauer der Äußerung aufgetragen. Damit beginnt jede Kurve mit der ersten Silbe bei dem Wert 0 und endet bei der letzten Silbe mit einem Wert kleiner als 100 Prozent (Startzeit der letzten Silbe liegt kurz vor dem Ende der Äußerung, es folgt aber noch die letzte Silbe). Durch diese Art der Darstellung wird der bei Sprechgesunden näherungsweise lineare Silben- bzw. Redefluß grafisch visualisiert, jede Redeflußstörung - z.B. durch das bei Patienten mit Sprechapraxiesyndrom typische Zögerphänomen meist unmittelbar vor dem Zielwort - tritt deutlich durch einen Knick im Kurvenverlauf hervor.
Das Menü DRUCKEN stellt diverse Varianten der Druckausgabe zur Verfügung, wie Farbdruck, Schwarz-Weiß-Druck oder Kopie in die Zwischenablage für die weitere Verwendung in klinischen Berichten usw.
Das Menü WIEDERGABE erlaubt dem Untersucher die akustische Wiedergabe einzelner Testsätze bei gleichzeitiger Betrachtung der ermittelten Kurvenverläufe.

Das zweite Auswertungsblatt gem. Abbildung 5.7 soll dem Untersucher einen Überblick verschaffen über die Ursachen eines ggf. beobachteten verlangsamten Redeflusses und auch über die Unterschiede in Abhängigkeit vom Zielwort. Hierzu werden 3 verschiedene Quotienten abgeleitet und dargestellt, das sind der

- PAUSENQUOTIENT: Anteil der Sprechpausen an der Gesamtdauer der Äußerung. Ein hoher Quotient deutet auf verlängerte Pausen hin;
- ITERATIONSQUOTIENT: Anteil der Iterationen an der Gesamtdauer der Äußerung. Ein hoher Quotient deutet auf Stottern hin, häufig beobachtbar vor dem Zielwort;
- ADDITIONSQUOTIENT: Anteil der Hinzufügungen an der Gesamtdauer der Äußerung. Ein hoher Quotient deutet hier auf häufige Fehlversuche oder Selbstkorrekturen hin.

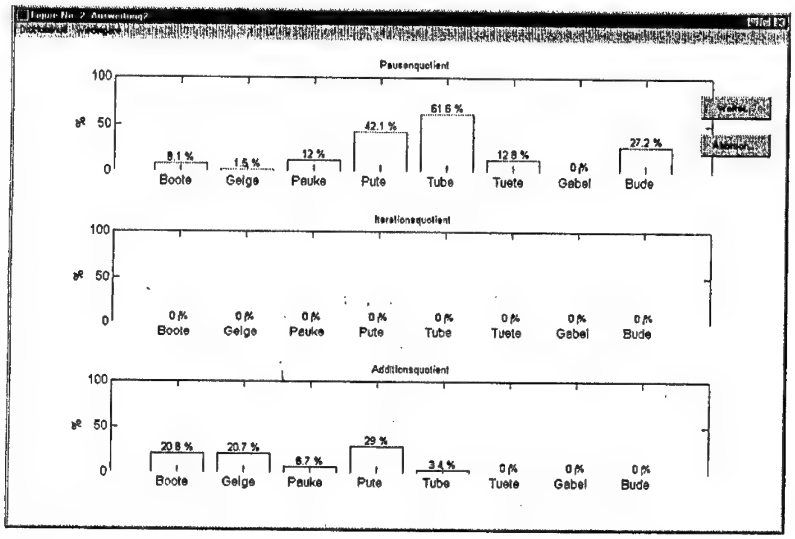

Abbildung 5.7.: Auswertung zum Modul SATZPRODUKTION, getrennt nach Zielwortkomplexität, Blatt 2. Die dargestellten Zeitquotienten geben Aufschluß darüber, wie effizient der Patient den Zeitraum der Nachsprechaufgabe zur Informationsübertragung nutzt.

Das dritte und letzte Auswertungsblatt gem. Abbildung 5.8 faßt die bisherigen Ergebnisse zusammen und bietet repräsentative Mittelwerte für die in Auswertung befindlichen Testsätze der gewählten Zielwortkomplexität. Es zeigt in der linken Hälfte eine Statistik über

- SILBENFEHLER: Iterationen (IT), Additionen (AD) bzw. Elisionen (E>),
- SEGMENTALE FEHLER: Phonetische Entstellungen (PE) oder Phonematische Paraphasien (PP),
- ZEITASPEKTE: Mittlere (Netto-)Silbendauer, mittlere Nettosilbenrate und mittlere Bruttosilbenrate.

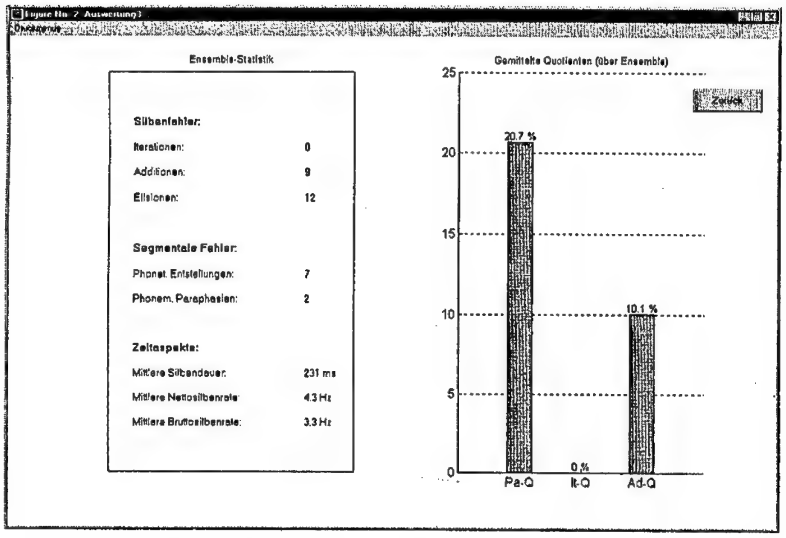

Abbildung 5.8.: Auswertung zum Modul SATZPRODUKTION, getrennt nach Zielwortkomplexität, Blatt 3.

Die beiden Fehlerzahlen ergeben sich durch Aufsummierung über alle betrachteten Testsätze. Die MITTLERE SILBENDAUER entspricht dem Gesamtmittelwert über die Silbendauern aller realisierter Silben der betrachteten Testsätze. Dieser Wert wurde bereits auf dem Auswertungsblatt 1 dargestellt und erscheint hier der Vollständigkeit halber.
Die MITTLERE NETTOSILBENRATE errechnet sich aus dem Kehrwert der MITTLEREN SILBENDAUER. Sie stellt damit das Frequenzanalogon zu diesem Zeitparameter dar und trägt die Dimension Hz.
Die MITTLERE BRUTTOSILBENRATE hingegen errechnet sich aus dem Kehrwert der Bruttosilbendauern, das sind die mittleren Zeitabstände zwischen den jeweils linken Silbengrenzen. Damit gehen auch die Sprechpausen in die weitere Bewertung ein. Diese Art der Unterscheidung ermöglicht in praxi u.a. die differenzierte Bewertung der Satzproduktion bei Patienten mit kurzen, abgehackten Silben. In diesem Falle werden sich die Brutto- und die Nettosilbenrate aufgrund der verlängerten Pausen erheblich unterscheiden, während die beiden Raten bei Sprechgesunden nur geringfügige Unterschiede

aufweisen.
In der rechten Bildhälfte sind die mittleren PAUSEN-, ITERATIONS- und ADDITIONSQUOTIENTEN der betrachteten Testsätze mit farbigen Balken visualisiert (Erläuterungen zu diesen Parametern siehe Seite 131).

5.2.3.2. Auswertung über alle Zielwortkomplexitäten

Während im vorhergehenden Abschnitt die Auswertung getrennt nach Zielworttypen durchgeführt wurde, erfolgt hier eine übergreifende Auswertung über alle 3 verschiedenen Zielworttypen hinweg. Damit sollen die Unterschiede und Abhängigkeiten in der Satzproduktion bei Zielworten unterschiedlicher artikulatorischer Komplexität sichtbar gemacht werden.
Hier werden nun nicht mehr repräsentative Parameterverläufe für jeden Testsatz einzeln, sondern nur noch die innerhalb einer Zielwortkomplexität gemittelten Verläufe sowie zusätzlich deren Gesamtmittelwert dargestellt. Der Aufruf dieser Auswertung ist nur möglich, wenn die Einzelauswertungen aller 3 Zielwortkomplexitäten zuvor bereits aufgerufen worden waren und damit sämtliche Teilergebnisse im System verfügbar sind.
Das erste Auswertungsblatt gem. Abbildung 5.9 widmet sich auch hier der grafischen Visualisierung von Tempo, Rhythmus und Redefluß, wobei die akustisch gewonnene Information mit der auditiven Information des Untersuchers geeignet verschränkt werden soll. Da jetzt sowohl zwei- als auch dreisilbige Zielworte vorkommen, wurde an dieser Stelle für beide Grafiken entlang der Abszisse ein festes, äquidistantes Raster mit 10 Zielsilben gewählt, wobei für die beiden Gruppen mit zweisilbigem Zielwort die nicht existierende Zielsilbe Nr. 7 (das wäre die dritte Silbe des Zielwortes) unberücksichtigt bleibt.
Im oberen Teil des Auswertungsblattes sind für jede der 3 Gruppen die mittleren Silbendauern in der Dimension ms über der entsprechenden Silbe aufgetragen. Diese Werte zeigen nun an, wie viel Zeit für die betreffende Silbe im Mittel aufgebracht wurde. Der resultierende Verlauf jeder Gruppe ist hier in einer eigenen Farbe dargestellt gemäß der unten rechts dargestellten Legende.
Das hier gezeigte Beispiel gem. Abbildung 5.9 entstammt ebenfalls einer Untersuchung an einem Patienten mit Sprechapraxie-Syndrom. Auch hier ist deutlich der Unterschied in den Silbendauern für die nicht akzentuierten Silben */te/* und */die/* gegenüber Silben mit Haupt- oder Nebenakzent (z.B. */kann/*) erkennbar. Die eingeblendete gestrichelte

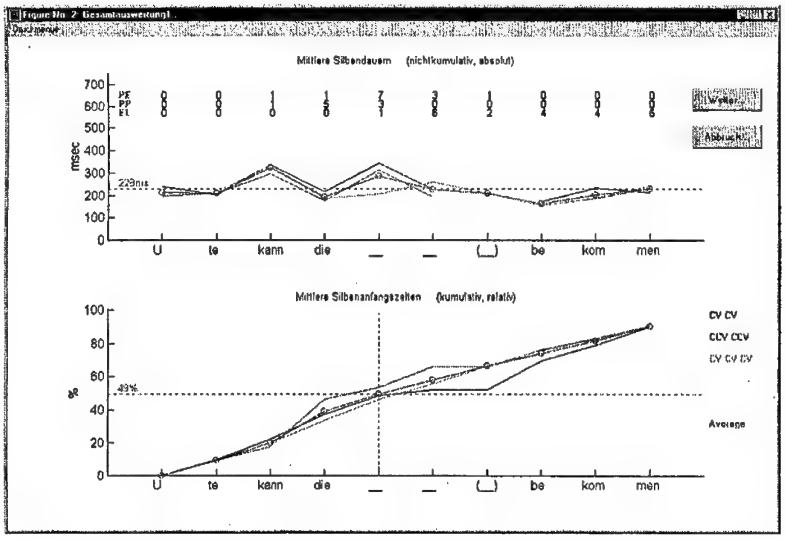

Abbildung 5.9.: Auswertung zum Modul SATZPRODUKTION. Gesamtdarstellung für alle 3 Zielwortkomplexitäten, Blatt 1

Linie zeigt den Durchschnittswert der Silbendauer über alle realisierten Silben aller 24 Testsätze (Grand Average), hier 228ms.

Die am oberen Rand dargestellte Tabelle gibt auch hier an, bei welcher Silbe und dort mit welcher absoluten Häufigkeit Phonetische Entstellungen (PE), Phonematische Paraphasien (PP) und Elisionen (EL) innerhalb der Sätze festgestellt wurden. Damit eröffnet sich dem Untersucher die Möglichkeit, quantifizierbare Aussagen bezüglich der Ursachen unflüssiger Satzproduktion zu treffen.

Im unteren Teil des Auswertungsblattes sind die Silbenstartzeitpunkte in ihrer relativen zeitlichen Lage bezogen auf die Gesamtdauer der Äußerung aufgetragen. Damit beginnt jede Kurve mit der ersten Silbe bei dem Wert 0 und endet bei der letzten Silbe mit einem Wert kleiner als 100 Prozent (Startzeit der letzten Silbe liegt kurz vor dem Ende der Äußerung). Durch diese Art der Darstellung wird der bei Sprechgesunden näherungsweise lineare Silben- bzw. Redefluß grafisch visualisiert, jede Flußstörung, z.B. durch die für Sprechapraktiker typischen Zögerphänomene (meist unmittelbar vor dem Zielwort),

tritt deutlich durch Knicke im Kurvenverlauf hervor. In dem gezeigten Beispiel eines Sprechapraktikers zögert der Patient vor den komplexen, zweisilbigen Zielwörtern mit Konsonantcluster besonders stark.
Das Menü DRUCKEN stellt diverse Varianten der Druckausgabe zur Verfügung, wie Farbdruck, Schwarz-Weiß-Druck oder Kopie in die Zwischenablage für die weitere Verwendung in klinischen Berichten usw.
Eine Wiedergabemöglichkeit wurde hier nicht implementiert, da nur gemittelte Kurven dargestellt werden und damit keine perzeptiv/visuelle Korrelation der Daten für den Untersucher sinnvoll erscheint.
Das zweite Auswertungsblatt gem. Abbildung 5.10 stellt dem Untersucher in Analogie zur vorhergehenden Auswertung gem. Abbildung 5.7 auf Seite 132 die drei beschriebenen Parameter

- PAUSENQUOTIENT,
- ITERATIONSQUOTIENT,
- ADDITIONSQUOTIENT

dar, jetzt jeweils gemittelt innerhalb aller Testsätze einer Zielwortkomplexität.
Im konkreten Beispiel ist der Additionsquotient für die Gruppe der komplexen, zweisilbigen Zielworte erhöht. Diese Information läßt im Zusammenhang mit den beobachteten Zögerphänomenen unmittelbar vor dem Zielwort den Schluß zu, daß an dieser Stelle jeweils unnötige Silben hinzugefügt, nicht aber iteriert werden. Die Ursache der verlangsamten Satzproduktion liegen hier also in einem artikulatorischen Suchverhalten des Patienten bei Konfrontation mit neuen und schwierig zu artikulierenden Wörtern begründet.

Das dritte und letzte Auswertungsblatt gem. Abbildung 5.11 zeigt in der linken Hälfte die Statistik über

- SILBENFEHLER: Iterationen (IT), Additionen (AD) bzw. Elisionen (E>),
- SEGMENTALE FEHLER: Phonetische Entstellungen (PE) oder Phonematische Paraphasien (PP),

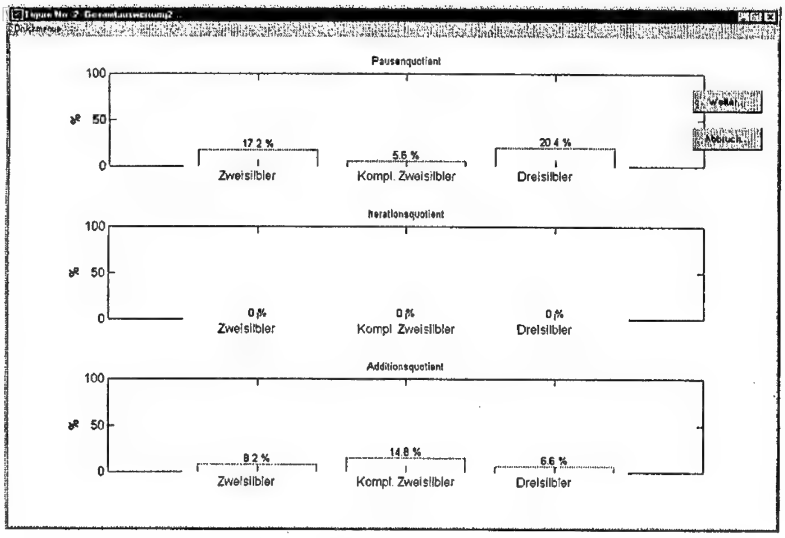

Abbildung 5.10.: Auswertung zum Modul SATZPRODUKTION. Gesamtdarstellung für alle 3 Zielwortkomplexitäten, Blatt 2

- ZEITASPEKTE: Mittlere (Netto-)Silbendauer, mittlere Nettosilbenrate und mittlere Bruttosilbenrate,

berechnet bzw. gemittelt über alle 24 Testsätze.
In der rechten Bildhälfte sind die über alle 3 Zielwortkomplexitäten gemittelten Pausen-, Iterations- und Additionsquotienten mit farbigen Balken visualisiert. Siehe hierzu auch die Interpretation auf Seite 131 und die vorhergehende Einzeldarstellung in Abbildung 5.7.

5.2.3.3. Automatischer Datenexport

Mit Aufruf einer Einzelauswertung gem. 5.2.3.1 auf Seite 129 und damit synchron zur Darstellung des jeweils ersten Auswertungsblattes gem. Abbildung 5.6 wird automatisch ein Datenfile mit allen relevanten Ergebnissen für die betrachtete Zielwortkomplexität

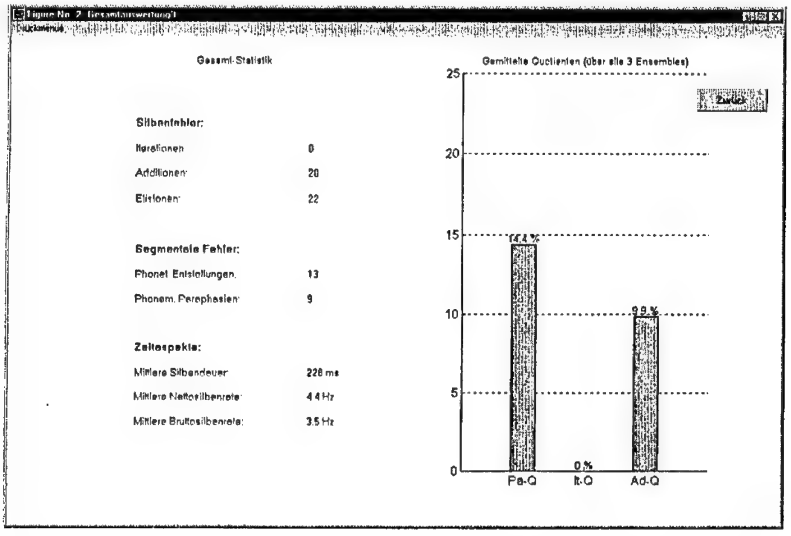

Abbildung 5.11.: Auswertung zum Modul SATZPRODUKTION. Gesamtdarstellung für alle 3 Zielwortkomplexitäten, Blatt 3

im entsprechenden Arbeitsverzeichnis angelegt. Das Datenfile ist ASCII-codiert und damit universell zur weiteren Nutzung oder Weiterverarbeitung mit alternativen Softwareprodukten einsetzbar. Der Filename ist an der Endung ..en1.txt / ..en2.txt / ..en3.txt identifizierbar je nach Zielworttyp.
Das Datenfile enthält ein Zahlenfeld in Matrixdarstellung, wobei jeder realisierte Testsatz einer separaten Matrixzeile zugeordnet ist. Jede Zeile enthält dabei in den ersten beiden Feldern (Spalten) folgende Informationen:

- Feld 01: Zielwortkomplexität, codiert gem. Tabelle 5.1,
- Feld 02: Zielwort, codiert gem. Tabelle 5.2.

Pro Zielsilbe der laufenden Nummer i (bei den Zielwortkomplexitäten 1 und 2 gilt $i = 1 \ldots 9$, bei Zielwortkomplexität 3 entsprechend $i = 1 \ldots 10$) wird nun jeweils ein 3-

elementiger Daten-Record angehängt. Damit ergeben sich für die Felder Nr. $3i \ldots 3i+2$ die folgenden Belegungen:

- Feld $3i+0$: Relative Startzeit der Zielsilbe Nr. i in Prozent,
- Feld $3i+1$: (Netto-)Silbendauer der Zielsilbe Nr. i in ms,
- Feld $3i+2$: Silbenattribut der Zielsilbe Nr. i, codiert gem. Tabelle 5.3.

Bei der Interpretation des Silbenattributes ist zu beachten, daß Zielsilben im Sinne der Definition keine wiederholten (Attribut IT) oder hinzugefügten Silben (Attribut AD) sein können. Siehe hierzu auch Seite 125. Zusätzlich zu den in Tabelle 5.3 aufgeführ-

Zielwortkomplexität	Code
Zweisilbige Wörter mit einfacher Konsonant-Vokal-Struktur	1
Zweisilbige Wörter mit Konsonant-Cluster	2
Dreisilbige Wörter	3

Tabelle 5.1.: Identifizierung der Zielwortkomplexität anhand der Codierung im Feld 01 des Datenfiles zum Modul SATZPRODUKTION.

Zielwort (Typ1)	Zielwort (Typ2)	Zielwort (Typ3)	Code
Boote	Kränze	Pakete	1
Geige	Platte	Tapete	2
Pauke	Knöpfe	Zigarre	3
Pute	Polster	Gitarre	4
Tube	Trümpfe	Trompete	5
Tüte	Klunker	Sandale	6
Gabel	Bretter	Tablette	7
Bude	Pflanze	Baracke	8

Tabelle 5.2.: Identifizierung des Zielwortes anhand der Codierung im Feld 02 des Datenfiles zum Modul SATZPRODUKTION.

ten Silbenattributen sind auch Doppelattribuierungen für Phonetische Entstellungen bei gleichzeitiger Phonematischer Paraphasie möglich. Die Codes ergeben sich in diesem Falle durch Kombination der beiden Einzelcodes für PE (an erster Stelle) und PP (an zweiter

Silbenattribut	Code
Reguläre Zielsilbe (ohne Attribut)	1
Elision E>	0
Verschmelzung VE	4
Phonetische Entstellung (nur Onset) PE1	6
Phonetische Entstellung (Onset und Rhyme) PE2	7
Phonematische Paraphasie (nur Onset) PP1	8
Phonematische Paraphasie (Onset und Rhyme) PP2	9

Tabelle 5.3.: Identifizierung des vergebenen Silbenattributs für die Silbe Nr. i anhand der Codierung im Feld $3i + 2$ des Datenfiles zum Modul SATZPRODUKTION.

Stelle), d.h., die Kombination aus PE1 und PP2 ergäbe dann beispielsweise den Code 69.
Die Matrix enthält noch 6 weitere Spalten bzw. Felder. Da die Testsätze mit Zielwortkomplexität 1 und 2 aus jeweils 9, die Testsätze mit Zielwortkomplexität 3 hingegen aus 10 Zielsilben bestehen, ergeben sich Einträge in die Felder 30...35 bzw. 33...38 nach folgender Konvention:

- Feld 30 bzw. 33: Additionsquotient,
- Feld 31 bzw. 34: Anzahl addierter Silben bzw. vergebener Attribute AD,
- Feld 32 bzw. 35: Iterationsquotient,
- Feld 33 bzw. 36: Anzahl iterierter Silben bzw. vergebener Attribute IT,
- Feld 34 bzw. 37: Pausenquotient,
- Feld 35 bzw. 38: Gesamtdauer der Äußerung in ms.

5.3. Technische Realisierung: Entwickelte Verfahren und Algorithmen

Die technische Aufgabenstellung liegt in dem Modul SATZPRODUKTION in der

- akustischen Analyse des Zeitmusters der Satzproduktion bei gleichzeitig
- auditiver Analyse nach phonetischen und phonematischen Fehlern.

Unter Zeitmuster soll im Folgenden die Information über die zeitliche Lage von Silbenkernen und Silbengrenzen verstanden werden. Der Lösungsansatz zur Analyse des Zeitmusters führt in das Gebiet der Spracherkennung, da hier unabhängig vom gewählten Erkennungsprozeß bereits in der Vorverarbeitungsstufe die Problematik der Halbsilben-, Silben- oder Einzelwortsegmentierung stets betrachtet werden muß ([Lea80]).
Die Erkennung von Silbengrenzen streift zusätzlich eine typische Problematik der Biosignalverarbeitung, nämlich die korrekte Detektion von Nutzsignalen (hier sind es Silben) in verrauschter Signalumgebung (hier im wesentlichen geprägt durch das Aufzeichnungssystem Soundkarte/Rechner). Von Interesse ist hier die Beantwortung der Frage, wann sich meßbare akustische Signalparameter deterministisch und signifikant ändern bei einer gleichzeitig möglichst hohen Robustheit gegenüber stochastischen Signalveränderungen z.B. durch Rauschpeaks, Geräuschartefakte usw.
Die Verschränkung der akustisch basierten Information bezüglich des Zeitmusters einerseits mit der auditiv basierten Information bezüglich segmentaler und suprasegmentaler Fehler andererseits stellen eine weitere Herausforderung an die Realisierung des Modules dar. Nur durch Verknüpfung und parallele Betrachtung beider Informationsebenen ist es möglich, eine quantifizierbare Aussage bezüglich der Ursachen und grundlegender Charakteristika einer gestörten Satzproduktion zu treffen.
Gleichzeitig muß vom System eine synchrone Überwachung bzw. Plausibilitätsprüfung dieser verknüpften Informationen realisiert werden, da weder Fehler in der automatischen Silbendetektion noch in der manuellen Eingabe auditiver Merkmale auf Grund der Gefahr von Fehlinterpretationen unerkannt bzw. unkorrigiert bleiben dürfen.

5.3.1. Detektion von Silbenkernen, Silbengrenzen und Sprechpausen

Der vorgestellte Segmentierungsalgorithmus fußt auf einer von RUSKE und BEHAM in [Rus92] vorgestellten Vorverarbeitung durch Berechnung der *modifizierten Lautheitsfunktion*. Durch Anwendung dieses Verfahrens entstehen Parameter-Zeitverläufe, die die menschliche Lautheitsempfindung über eine Modellierung der physiologischen Vorgänge im Gehör nachbilden und damit für die algorithmische Silbensegmentierung eine optimale Detektionsgrundlage bereitstellen.

5.3.1.1. Psychoakustische Grundlagen des Hörens

Zunächst muß an dieser Stelle erläuternd auf die psychoakustischen Grundlagen des Hörens eingegangen werden.
Die menschliche Sprachwahrnehmung wertet Schallsignale aus, die auf unser Gehörorgan treffen. Hier werden akustische Schallwellen in Wanderwellen entlang der Basilarmembran in der Gehörschnecke umgewandelt. Da der Ort der maximalen Auslenkung auf der Basilarmembran von der Frequenz abhängig ist, findet dort in erster Linie eine spektrale Zerlegung des Schallsignales in seine Frequenzanteile statt. Die Sinneszellen (Haarzellen) sowie die neuronale Verarbeitung werden dafür verantwortlich gemacht, daß die Frequenzselektivität gegenüber der begrenzten Wirksamkeit der reinen mechanischen Auslenkung wesentlich verbessert wird ([Nil78]).
HELMHOLTZ erarbeitete bereits 1863 die sogenannte Ortstheorie, nach der ein Ton im Innenohr eine Reizverteilung längs des CORTIschen Organs hervorruft, welche an einer mit der Tonfrequenz eindeutig zusammenhängenden Stelle ein Maximum hat ([Hel1863]). Nach diesem Prinzip des adäquaten Reizes ([Mue1840]) gehört zu jeder Stelle des CORTIschen Organs eine Tonhöhenempfindung (*Tonotopie*).
Diese Theorie hat sich im Laufe der Zeit als unzulänglich erwiesen. Es wurden Tonhöhenphänomene bekannt, welche die Ortstheorie nicht erklären kann. Ein im Bereich der alltäglichen Erfahrung liegendes Beispiel hierfür ist die Wahrnehmung der menschlichen Stimme am Telefon, obgleich die Grundfrequenzen der meisten stimmhaften Klänge unterhalb von 300 Hz liegen und damit nicht übertragen werden. Trotzdem nimmt das menschliche Gehör die Stimme des Gesprächpartners in ihrer natürlichen Tonlage wahr. Die Arbeiten von TERHARDT erweitern und ergänzen die klassische Vorstellung der

Tonhöhenwahrnehmung um weitere Annahmen, die sich auf Beobachtungsergebnisse stützen ([Ter72]). Er unterschied zwischen

- *reinen* Tönen, deren Tonhöhe sich nach der HELMHOLTZschen Ortstheorie verständlich machen lässt, und
- *komplexen* Tönen, deren Klangtonhöhe als ein sekundäres Wahrnehmungsattribut aufgefaßt werden kann, welches aus den Einzeltönen dominanter Harmonischer abgeleitet ist.

5.3.1.2. Funktionsmodell der Lautheitsempfindung

Aufbauend auf den Erkenntnissen von HELMHOLTZ und TERHARDT stellte VOGEL ein Funktionsmodell der Lautheit vor, mit dessen Hilfe sich die Verteilung der Lautheit über der Frequenz bzw. der psychoakustischen Empfindungsgröße Tonheit nachbilden läßt ([Vog75]).
Grundlegende Schritte hierfür sind die

- Transformation der Frequenz in die psychoakustische Empfindungsgröße *Tonheit* sowie die
- Transformation des Schalldruckes in die psychoakustische Empfindungsgröße *Lautheit*.

Als Ausgangspunkt hierfür dient die spektrale Zerlegung des gesamten hörbaren Bereiches in 24 Frequenzgruppen nach ZWICKER ([Zwi61]). Das Ergebnis ist eine nichtlineare Frequenzskala, die als Tonheit z bezeichnet wird.
Das Funktionsmodell der Lautheit nach VOGEL ([Vog75]) berücksichtigt zusätzlich die zeitlichen Integrationseffekte der Vor- und Nachverdeckung ([Zwi77]). Es erlaubt somit die Nachbildung des aus der Psychoakustik bekannten Phänomens, daß leise Töne von vorausgegangenen lauten Tönen verdeckt und nicht wahrgenommen werden.
RUSKE UND BEHAM schlagen in [Rus92] vor, auf diese Modellierung der zeitlichen Effekte zu verzichten und dadurch die weitere Berechnung erheblich zu vereinfachen. Damit wird nur noch die spektrale Auflösung bei der Gewinnung der Erregungsverteilung berücksichtigt. Zusätzlich werden hier handhabbare und relativ leicht implementierbare

5. Realisierung des Moduls SATZPRODUKTION

Berechnungsverfahren beschrieben, welche bei der Realisierung des MODIAS-Moduls SATZPRODUKTION weitgehend umgesetzt wurden.

Die Berechnung der gesuchten und von RUSKE und BEHAM vorgeschlagenen Parameter-Zeitverläufe der Lautheit $N(t)$ und der modifizierten Lautheit $N_m(t)$ läßt sich funktional in folgende Schritte untergliedern:

- Schätzung von Kurzzeit-Leistungsdichtespektren bzw. Berechnung der physikalischen Leistungsdichteverteilung $P(f,t)$,
- Aufteilung des Leistungsdichtespektrums $P(f,t)$ in i=22 gehörrichtige Frequenzgruppen z_i,
- Berechnung der psychoakustischen Erregungsverteilung $E(z_i,t)$ durch Nachbildung spektraler Integrationseffekte,
- Umrechnung von der Erregungsverteilung $E(z_i,t)$ auf die Lautheitsverteilung $N'(z_i,t)$,
- Berechnung der Lautheit $N(t)$ und der modifizierten Lautheit $N_m(t)$ durch Linearkombinationen von Elementen der Lautheitsverteilung $N'(z_i,t)$.

Schätzung von Kurzzeit-Leistungsdichtespektren und Frequenz-Tonheitstransformation: Das hier umgesetzte Verfahren zur Schätzung des Leistungsdichtespektrums basiert auf Anwendung der diskreten Fourier-Transformation (berechnet nach dem Fast-Fourier-Transformationsalgorithmus FFT) und anschließender Quadrierung der resultierenden spektralen Amplitudendichte. Wie bei jedem Kurzzeit-Analyseverfahren muß dabei zunächst durch Anwendung einer geeigneten Fensterfunktion ein zeitlich begrenztes Signalsegment entnommen werden, wodurch die Schätzung des Leistungsdichtespektrums verfälscht wird. Dieser Einfluß wird durch eine geeignete Skalierung kompensiert.
Das in MODIAS realisierte Berechnungsverfahren sieht in einem ersten Schritt die Aufteilung der gesamten Aufzeichnung (Samplingrate 22050 Hz / Auflösung 16 Bit) in Segmente zu je 512 Samples bei einer zeitlichen Überlappung von 50 Prozent bzw. 256 Samples vor. Für jedes dieser Segmente wird ein Leistungsdichtespektrum nach der

dargestellten Berechnungsmethode unter Nutzung von Hamming-Fenstern ermittelt.
Die Frequenzstützstellen der errechneten Leistungsdichtespektren sind zunächst noch äquidistant auf einer linearen Frequenzskala verteilt. Abbildung 5.12 zeigt ein typisches Leistungsdichtespektrum am Beispiel eines Vokals.

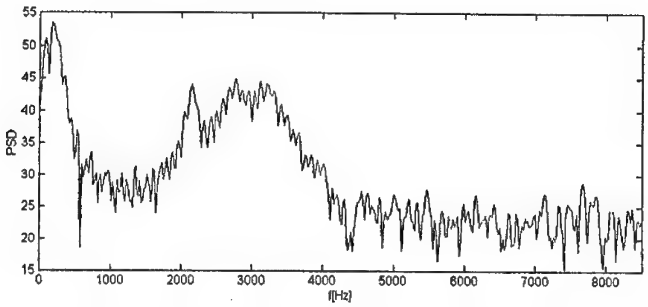

Abbildung 5.12.: Typisches Leistungsdichtespektrum eines Signalsegmentes des Vokals *i* nach entsprechender Berechnung der Fast-Fourier-Transformation. Die Frequenzachse ist linear skaliert, die Frequenzstützstellen des Spektrums sind äquidistant entlang der Frequenzskala verteilt.

Die geforderte Frequenz-Tonheitstransformation erfolgt durch Umskalierung der Frequenzachse f in die Tonheitsache z. Der funktionale Zusammenhang zwischen Frequenz und Tonheit ist z.B. in [Zwi82] angegeben, siehe hierzu auch Formel 7.3 auf Seite 224. Durch diese Umskalierung wird das Frequenzspektrum nichtlinear entlang der Frequenzachse gespreizt, siehe hierzu Abbildung 5.13. Wie auch beim menschlichen Gehör ist dadurch die Frequenzauflösung bei niedrigen Frequenzen erhöht und bei hohen Frequenzen vermindert.
Im nächsten Schritt erfolgt die Einteilung der Tonheitsachse in $i = 1 \ldots 22$ gehörrichtige Frequenzbänder bzw. Kanäle. Der Tonheitsbereich $z = i - 1 \ldots i$ bildet dabei jeweils das Frequenzband i. Die Tabelle 7.1 auf Seite 254 stellt alle Bänder mit ihren gem. Formel 7.3 definierten unteren und oberen Frequenzgrenzen orientierend dar.
Durch Kanalbildung entstehen damit aus jedem der ursprünglichen Leistungsdichtespektren 22 Frequenzgruppen-Spektren, deren spektrale Verteilung jedoch noch auf die tatsächliche Erregungsverteilung des menschlichen Gehörs umgerechnet werden muß.

Abbildung 5.13 zeigt das umskalierte Leistungsdichtespektrum (vgl. Abbildung 5.12) mit den resultierenden Kanalgrenzen.

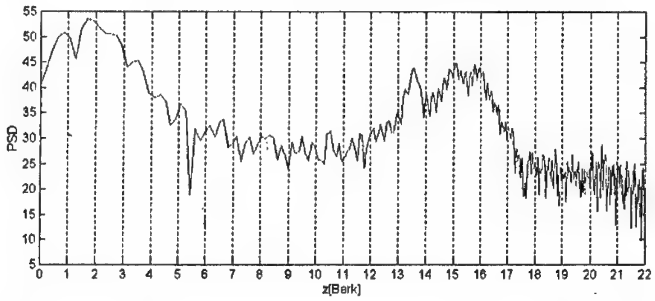

Abbildung 5.13.: Gegenüber Abbildung 5.12 umskaliertes Leistungsdichtespektrum eines Signalsegmentes mit nichtlinearer Tonheitsskala und damit nicht mehr äquidistanter Verteilung der Frequenzstützstellen. Zusätzlich sind die Kanalgrenzen der Bänder 1 ... 22 eingezeichnet.

Berechnung der psychoakustischen Erregungsverteilung und Erregungs-Lautheits-Transformation: Die Erregung E ist eine psychoakustische Empfindungsgröße; sie wird meist logarithmiert als Erregungspegel L_E angegeben ([Zwi67]). Die Verteilung des Erregungspegels L_E über der Tonheit z läßt sich durch Faltung des Frequenzgruppen-Spektrums mit einer Schablone berechnen, die der Verteilung der spektralen Energie einer Frequenzgruppe entspricht. In erster Näherung kann die Form der Schablone als dreieckig angenommen werden, wobei die Flanke zu tieferen Frequenzen etwa 25dB/Bark und der Abfall zu höheren Frequenzen etwa -10dB/Bark beträgt (Abbildung 5.14).
Diese Form der Schablone ist unabhängig von der Lage auf der Tonheitsachse, wobei die vorhandene leichte Abhängigkeit vom Schallpegel bei Anwendungen in der automatischen Spracherkennung meist vernachlässigt wird ([Rus92]).
Zur konkreten Berechnung wird als repräsentative Schablone der experimentell ermittelte Erregungspegel $L_{E_{1kHz}}(z)$ herangezogen, der sich bei einer Anregung des menschlichen Gehörs mit Schmalbandrauschen der Mittenfrequenz 1000 Hz und einem Schallpegel von 60 dB einstellt ([Zwi82]).

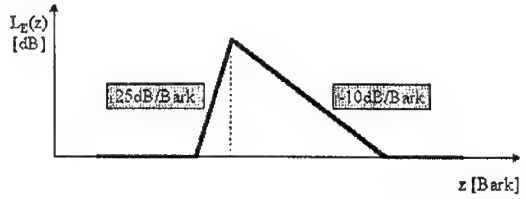

Abbildung 5.14.: Verwendete Schablone zur Modellierung der spektralen Maskierungseffekte des menschlichen Gehörs gem. [Zwi82]

Die Erregungsverteilung $E(z_i)$ ergibt sich durch Verschiebung an die Stelle z_i der Tonheitsachse.
Die so berechnete Verteilung des Erregungspegels $L_E(z)$ repräsentiert damit insgesamt sowohl die hydromechanische als auch die neuronale Vorverarbeitung des Gehörs.
Da die Faltungsoperation für alle 22 Frequenzgruppenspektren durchgeführt werden muß, ist hier die Einführung einer Schablonen-Matrix M sinnvoll. Beträgt die Länge der FFT N=512, dann ist aus Symmetriegründen die Breite der Leistungsdichtespektren gleich $N/2$+1 bzw. 257. Die Spalten j der Schablonen-Matrix M sind jeweils den entsprechenden Komponenten des Leistungsdichtespektrums $P(f_j)$ an den diskreten Frequenzstützstellen $f_j = j\Delta f$ mit $0 \leq j < 256$ zugeordnet.
Eine Zeile i der Matrix beschreibt die Form der Schablone zur Berechnung eines Wertes der Erregungsverteilung mit

$$E_i = E(z_i) = \sum_{j=0}^{N/2} M_{ij} P(f_j) \tag{5.1}$$

Die Zeilen der Schablonen-Matrix stellen damit eine Schar von Schablonen dar, die jeweils auf der Tonheitsachse verschoben sind. Zu beachten ist hierbei, daß die Schablonen für die Berechnungsweise nach Gleichung 5.1 bezüglich Amplitude und Frequenz linear skaliert sein müssen.

Abbildung 5.15 zeigt beispielhaft die Matrixzeilen bzw. Schablonen für die Bark-Bänder 10, 15, 20 und 22 (von links nach rechts) auf linearer Frequenz- und Amplitudenskala.

Nach Anwendung der Faltungsoperation bzw. der Matrixmultiplikation gem. Gleichung 5.1 ist also aus jedem ursprünglich frequenzskalierten Leistungsdichtespektrum

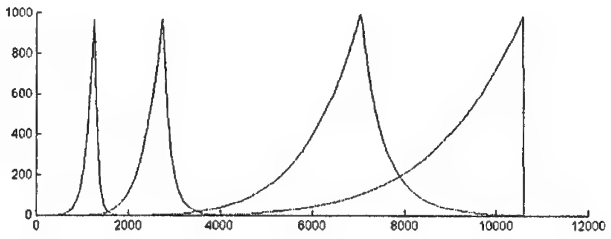

Abbildung 5.15.: Schablonen, die für die Transformation von Frequenzgruppenspektren auf Frequenzgruppen-Erregungsverteilungen verwendet werden und damit die spektralen Maskierungseffekte des menschlichen Gehörs nachbilden. Hier dargestellt die Schablonen für die Bark-Bänder 10,15,20 und 22, jedoch gegenüber Abbildung 5.14 bezüglich beider Achsen linear skaliert.

der Länge 256 jeweils eine tonheitsskalierte Erregungsverteilung der Länge 22 entstanden. Die Berechnung der Kanalgrenzen ist auf Seite 224 ausführlich dargestellt. Abbildung 5.16 zeigt die resultierende Erregungsverteilung (vgl. auch mit Abbildung 5.13).

Die Lautheitsempfindung eines Schallsignales hängt sowohl vom Schallpegel als auch von dessen spektraler Verteilung ab. Grundsätzlich kann die Lautheit N als Integral über die spezifische Lautheit $N'(z)$ entlang der gesamten Tonheitsachse berechnet werden ([Zwi67]). Im Falle einer wie hier diskreten Erregungsverteilung wird die Lautheit N als Summe der spezifischen Lautheitskomponenten N'_i über alle i Frequenzgruppen bestimmt.

Wird der Einfluß der Ruhehörschwelle vernachlässigt, läßt sich die spezifische Lautheitskomponente N'_i aus der Erregung E_i durch folgenden Zusammenhang berechnen:

$$N'_i \sim E_i^{\ 0.23} \tag{5.2}$$

Durch Anwendung des Potenzgesetzes entstehen somit über eine einfache Umskalierung der Amplitudenachse aus den 22 Komponenten der Erregungsverteilung E_i die 22 spezifischen Lautheitskomponenten N'_i.

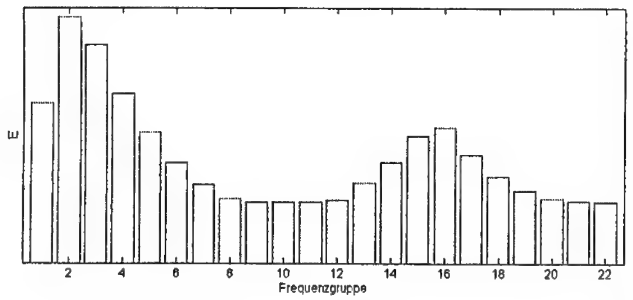

Abbildung 5.16.: Erregungsverteilung eines Signalsegmentes über die 22 gehörrichtigen Tonheitskanäle. Durch die beschriebene Faltungsoperation jedes Frequenzgruppenspektrums gem. Abbildung 5.12 mit der Schablone gem. Abbildung 5.14 bzw. 5.15 entsteht pro Kanal ein diskreter Wert, der mit der psychoakustischen Erregung des menschlichen Gehörs korrespondiert.

Werden alle diese 22 spezifischen Lautheitskomponenten aufsummiert, entsteht die (Gesamt-)Lautheit N des betrachteten Signalsegmentes.

5.3.1.3. Lautheitsfunktion und modifizierte Lautheitsfunktion

Bedingt durch die Zeitvarianz des Sprachschalles ist die spezifische Lautheit $N_i'(t)$ wie auch die zugehörige Lautheit $N(t)$ eine Funktion der Zeit. Der gesuchte zeitliche Lautheitsverlauf bzw. die von RUSKE und BEHAM vorgeschlagene Lautheitsfunktion $N(t)$ läßt sich damit annähern als:

$$N(t) = \sum_{i=1}^{24} N_i'(t) \tag{5.3}$$

Die modifizierte Lautheitsfunktion entsteht als Differenz der beiden Teillautheiten in den Kanälen $i = 3 \ldots 15$ und $i = 20 \ldots 22$ gem. [Rus94]:

$$N_m(t) = \sum_{i=3}^{15} N_i'(t) - \sum_{i=20}^{22} N_i'(t) \tag{5.4}$$

Sie bildet die Ausgangsbasis für die Anzeige der Silbenkerne, da die modifizierte Lautheitsfunktion aufgrund der Differenzbildung große Werte annimmt, wenn das Spektrum des Sprachsignales im mittleren Frequenzbereich verteilt ist. Dies ist vor allem bei Vokalen der Fall. Energiereiche Frikativ- bzw. Plosivlaute werden durch die negative Gewichtung der oberen Frequenzkanäle gedämpft.
Während die Lautheitsfunktion in der dargestellten Form direkt Anwendung findet und besonders geeignet für die Detektion von Silbengrenzen ist, wird die modifizierte Lautheitsfunktion $N_m(t)$ zusätzlich mit Hilfe eines Tiefpaßfilters geglättet. Nach Glättung können die vorhandenen Maxima dieser Funktion dann unmittelbar als Kandidaten für die Silbenkerne herangezogen werden.

Die Auswahl und Dimensionierung des Glättungsfilters stützt sich auf experimentelle Untersuchungen von GEYWITZ gem. [Gey84]. Es kann auf einfache Weise unter Verwendung eines elementaren nichtkausalen Tiefpaßfilters mit rechteckiger Impulsantwort realisiert werden; das Ausgangssignal $N_m(i)$ zu einem Zeitpunkt i berechnet sich aus dem Eingangssignal $\hat{N}_m(i)$ zu:

$$N_m(i) = 1/3(\hat{N}_m(i-1) + \hat{N}_m(i) + \hat{N}_m(i+1)) \tag{5.5}$$

Wird dieses Filter mehrmals in Reihe geschaltet, so entsteht eine Impulsantwort, die mit wachsendem Wiederholungsfaktor k schließlich die Form einer Gauß-Funktion annimmt. Die Anzahl der Kaskadierung sollte an die Sprechgeschwindigkeit des Patienten angepaßt werden. Die Untersuchungen von GEYWITZ an Normalsprechern zeigen ein optimales Segmentierungsverhalten eines entsprechenden Silbenerkenners, wenn die modifizierte Lautheitsfunktion mit einer Wiederholungszahl von $k = 7$ geglättet wurde. Dieses Filter wurde daher für unauffällige Sprechtempi mit $k = 7$ in MODIAS implementiert. Zusätzlich kann hier die Wiederholungszahl an das subjektiv empfundene Sprechtempo bei Bedarf angepaßt werden.
Abbildung 5.17 zeigt die nach dem beschriebenen Verfahren berechnete Lautheitsfunktion und die modifizierte Lautheitsfunktion für den Satz *Ute kann die Platte bekommen.*

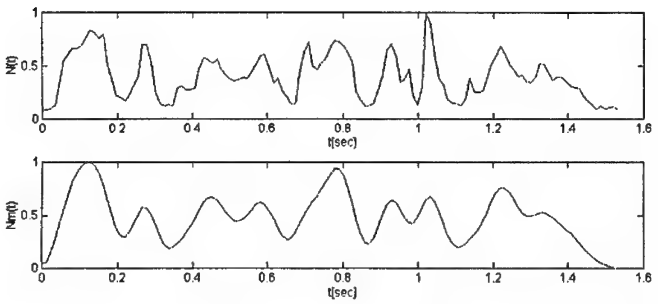

Abbildung 5.17.: Lautheitsfunktion (oben) und modifizierte Lautheitsfunktion (unten) des Satzes *Ute kann die Platte bekommen.* Letztere eignet sich aufgrund ihrer Sensitivität für Vokale hervorragend für die Detektion von Silbenkernen.

5.3.1.4. Optimale Detektion von Silbenkernen und Silbengrenzen

Zur optimalen Detektion der Silbenkerne eignet sich die modifizierte Lautheitsfunktion, also die Auswertung der spektralen Information der segmentweisen Lautheitsverteilungen ([Rus94]). Ein ähnliches Verfahren zur Anzeige der Silbenkerne wurde auch in [Rie81] vorgeschlagen.
Bei der Realisierung des MODIAS-Moduls SATZPRODUKTION findet hierzu ein Algorithmus Anwendung, der

- die Zeitpunkte sämtlicher lokalen Maxima im Verlauf der modifizierten Lautheitsfunktion detektiert,
- die zugehörigen modifizierten Lautheiten am Ort dieser Maxima mit einem empirisch ermittelten Schwellwert vergleicht (Plausibilitätsprüfung) und damit
- Kandidaten für Silbenkerne bereitstellt.

Werden in einem ersten Durchlauf mehr als 15 Kandidaten für die Silbenkerne detektiert (empirischer Wert), erfolgt automatisch eine Anpassung der sprechgeschwindigkeitsabhängigen Filterung und damit eine stärkere Glättung der modifizierten

Lautheitsfunktion.
Hierbei reduziert sich erfahrungsgemäß die Anzahl der Kandidaten deutlich. Während bei einem Sprechgesunden die Anzahl der tatsächlich realisierten Silben in fast allen Fällen mit 9 bzw. 10 auf Grund des angebotenen Satzmateriales vorhersagbar ist, muß bei Patienten mit neurogenen Störungen mit einer erhöhten (Iterationen, Hinzufügungen) oder verminderten Anzahl (Elisionen) gerechnet werden.

Die Minima im korrespondierenden Lautheitsverlauf zeigen mögliche Stellen für Silbengrenzen an ([Rus94]). Eine ähnliche Silbenabgrenzung wird auch in ([Merm75],[Hun80]) vorgenommen.
An dieser Stelle wurde auf Basis der Lautheitsfunktion ein Algorithmus implementiert, der

- die Zeitpunkte sämtlicher lokaler Minima im Verlauf der Lautheitsfunktion detektiert,
- die zugehörigen Lautheiten am Ort dieser Minima mit einem statistisch ermittelten Schwellwert vergleicht (er ergibt sich aus dem inherenten Grundrauschen) und damit
- sowohl Kandidaten für Silbengrenzen bereitstellt als auch auftretende Sprechpausen erkennt.

Bei der Detektion von Silbengrenzen ist an dieser Stelle zu beachten, daß zwischen zwei benachbarten Silben eine Sprechpause auftreten kann, diese aber nicht immer auftreten muß. Üblicherweise werden Silbenübergänge bei normaler Satzproduktion soweit verschliffen, daß kein signifikanter Abfall der Lautheitsfunktion und damit keine Pause festgestellt werden kann.
In der konkreten Anwendung sind die Sprechpausen in ihrer zeitlichen Dauer diagnostisch relevant, sie müssen möglichst exakt von den Silben (also Segmenten mit artikulatorischer Aktivität) unterschieden werden. Daher wurde das Verfahren gem. [Rus94] entsprechend modifiziert.

Der realisierte Segmentierungsalgorithmus berechnet zunächst aus der zu jeder Aufzeichnungssitzung erstellten Stilleaufzeichnung die zugehörige Lautheitsfunktion.

Aus diesem Verlauf wird anschließend statistisch ein Schwellenwert ermittelt, dessen Überschreitung mit hinreichender Sicherheit die Übergänge von den Sprechpausen auf die Bereiche artikulatorischer Aktivität anzeigt. Zur Ermittlung des Schwellenwertes wird für die Stilleaufzeichnung

- die Wahrscheinlichkeitsdichtefunktion des Lautheitsverlaufes über ein Histogrammverfahren geschätzt,
- die Wahrscheinlichkeitsverteilungsfunktion durch Integration aus der Wahrscheinlichkeitsdichtefunktion berechnet und
- die Lautheit detektiert, bei der die streng monotone Wahrscheinlichkeitsverteilungsfunktion den Wert 0.9 annimmt.

Damit definiert der gesuchte Schwellwert im Umkehrschluß diejenige Lautheit, die nur mit einer Wahrscheinlichkeit von 10 Prozent von dem systemspezifischen Grundrauschen überschritten wird und damit mit einer Signifikanz von 90 Prozent von artikulatorischer Aktivität herrührt. Abbildung 5.18 zeigt die Wahrscheinlichkeitsverteilungsfunktion der Lautheit für eine typische Stilleaufzeichnung.

Die Minima der Lautheitsfunktion wurden bereits berechnet, sie stehen als mögliche Kandidaten für die Silbengrenzen zur Verfügung. Der Segmentierungsalgorithmus unterscheidet nun die beiden Fälle, daß jeweils zwischen 2 aufeinanderfolgenden Silbenkernen

- genau 1 Silbengrenzenkandidat zu liegen kommt, der eine Lautheit oberhalb des zugehörigen Schwellwertes besitzt. Damit definiert diese Silbengrenze das Ende der vorhergehenden und zugleich den Beginn der nachfolgenden Silbe;
- mehr als 1 Silbengrenzenkandidat zu liegen kommt, wobei alle eine Lautheit oberhalb des zugehörigen Schwellwertes besitzen. Gewählt wird aus dieser Menge der Silbengrenzenkandidat, der die geringste Lautheit besitzt (absolutes Minimum). Auch hier definiert die so detektierte Silbengrenze das Ende der vorhergehenden und zugleich den Beginn der nachfolgenden Silbe;

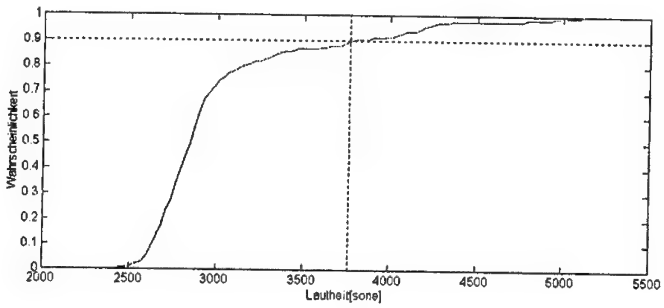

Abbildung 5.18.: Typische Wahrscheinlichkeitsverteilungsfunktion einer Stilleaufzeichnung. Gestrichelt eingezeichnet ist die Lautheit, bei der die Funktion den Wert 0.9 annimmt. Dieser Wert wird für jede Aufzeichnungssitzung aus der im Untersuchungsprotokoll enthaltenen Stilleaufzeichnung neu ermittelt. Er definiert für sämtliche Folgeaufzeichnungen dieser Sitzung den Schwellwert für die Detektion von Silbengrenzen.

- ein oder mehrere Silbengrenzenkandidaten zu liegen kommen, wobei mindestens einer von diesen eine Lautheit unterhalb des zugehörigen Schwellwertes besitzt. Der Algorithmus detektiert die Durchtrittspunkte der Lautheitsfunktion durch den Schwellwert und definiert somit für den Bereich der Unterschreitung eine Sprechpause. Die linke Grenze dieser Sprechpause definiert entsprechend das Ende der vorhergehenden Silbe, die rechte Grenze den Beginn der nachfolgenden Silbe.

Abbildung 5.19 stellt diese Segmentierungslogik an einem Beispiel erläuternd dar.

5.3.2. Akustische Analyse zeitlicher Aspekte der Sprachproduktion

Nach Abschluß der automatischen - ggf. auch manuell korrigierten - Segmentierung sind die Orte der Silbenkerne und der Silbengrenzen festgelegt. Das MODIAS-System ist auf akustischem Wege prinzipiell nur in der Lage, akustische Ereignisse mit silbischem Charakter zu detektieren. Die Differenzierung zwischen störendem Artefakt und einer tatsächlich produzierten Silbe wäre für Normsprecher über ein entsprechendes Erkennungsver-

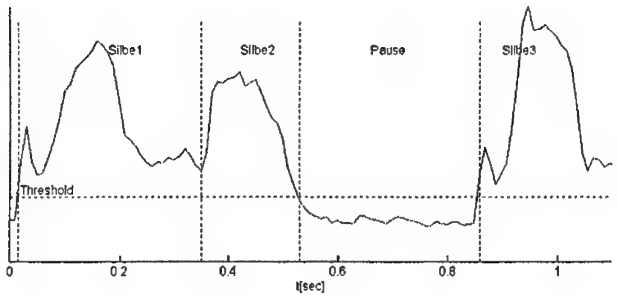

Abbildung 5.19.: Verlauf der Lautheitsfunktion für die ersten 3 Silben des Testsatzes *Ute kann die Tüte bekommen*. Der Schwellwert (Threshold) legt dabei den Pausenabschnitt zwischen */U/te/* und */kann/* fest, während zwischen den Silben */U/* und */te/* nur eine Silbengrenze ohne Pause erkannt wird.

fahren mit geringer Fehlklassifikationsrate möglich, jedoch kann diese Differenzierung bei Patienten mit neurogenen Sprechstörungen und den damit verbundenen akustischen Variationen wohl nur durch den Menschen selbst erfolgen. Die akustische Analyse zeitlicher Aspekte der Sprachproduktion kann bei neurogenen Sprechstörungen erst in Verbindung mit der auditiven Analyse segmentaler Fehler sinnvoll interpretiert werden.
Für die folgende Betrachtung liege eine unauffällige Satzproduktion des Testsatzes *Ute kann die Platte bekommen* vor; die Segmentierung in N Silben (hier: 9) sei korrekt erfolgt. MODIAS erstellt für jeden Satz eine Silbenmatrix S, deren Zeilen mit den Silben $i = 1 \ldots N$ korrespondieren. Die 3 Spalten beinhalten die detektierten Zeitpunkte für Silbenbeginn ($t_{i,1}$), Silbenkern ($t_{i,2}$) und Silbenende ($t_{i,3}$). Per definitionem beginnt jede Aufzeichnung mit der ersten Silbe bei $t = 0$, d.h., für alle weiteren Betrachtungen gilt $t_{1,1} = 0$.

$$S = \begin{pmatrix} t_{1,1} & t_{1,2} & t_{1,3} \\ t_{2,1} & t_{2,2} & t_{2,3} \\ \ldots\ldots\ldots\ldots\ldots \\ t_{N,1} & t_{N,2} & t_{N,3} \end{pmatrix} = \begin{pmatrix} 0 & t_{1,2} & t_{1,3} \\ t_{2,1} & t_{2,2} & t_{2,3} \\ \ldots\ldots\ldots\ldots\ldots \\ t_{N,1} & t_{N,2} & t_{N,3} \end{pmatrix} \tag{5.6}$$

Die in der Auswertung des MODIAS-Modules dargestellten (Netto-)SILBENDAUERN ergeben sich durch silbenweise Differenzbildung zwischen den Zeitpunkten von Silbenende und Silbenanfang bzw. $Tnetto_i = t_{i,3} - t_{i,1}$.
Damit entsteht pro Testsatz aus den N Nettosilbendauern ein Vektor $Tnetto$ gemäß folgender Berechnungsweise:

$$Tnetto = \begin{pmatrix} Tnetto_1 \\ Tnetto_2 \\ \cdots\cdots\cdots \\ Tnetto_N \end{pmatrix} = \begin{pmatrix} t_{1,3} - t_{1,1} \\ t_{2,3} - t_{2,1} \\ \cdots\cdots\cdots \\ t_{N,3} - t_{N,1} \end{pmatrix} = \begin{pmatrix} t_{1,3} \\ t_{2,3} - t_{2,1} \\ \cdots\cdots\cdots \\ t_{N,3} - t_{N,1} \end{pmatrix} \tag{5.7}$$

Die grafische Darstellung gem. Abbildung 5.6 auf Seite 130 bzw. Abbildung 5.9 auf Seite 135 (jeweils obere Bildhälfte) zeigt die gem. Gleichung 5.7 errechneten Nettosilbendauern, aufgetragen jeweils über der laufenden Nummer der korrespondierenden Silbe.

Nicht dargestellt, aber dennoch bei weiteren Berechnungen verwendet werden zusätzlich auch die Bruttosilbendauern. Sie ergeben sich durch Differenzbildung jeweils zwischen den Zeitpunkten aufeinanderfolgender Silbenbeginne bzw. $Tbrutto_i = t_{i+1,1} - t_{i,1}$.
Damit entsteht pro Testsatz aus den $N-1$ Bruttosilbendauern ein Vektor $Tbrutto$ mit

$$Tbrutto = \begin{pmatrix} Tbrutto_1 \\ Tbrutto_2 \\ \cdots\cdots\cdots\cdots \\ Tbrutto_{N-1} \end{pmatrix} = \begin{pmatrix} t_{2,1} - t_{1,1} \\ t_{3,1} - t_{2,1} \\ \cdots\cdots\cdots\cdots \\ t_{N,1} - t_{N-1,1} \end{pmatrix} = \begin{pmatrix} t_{2,1} \\ t_{3,1} - t_{2,1} \\ \cdots\cdots\cdots\cdots \\ t_{N,1} - t_{N-1,1} \end{pmatrix} \tag{5.8}$$

Die Bewertung des Silbenflusses erfolgt über die grafische Darstellung der relativen Zeitpunkte des Beginns jeder der 9 bzw. 10 Silben des Testsatzes (Silbenanfangszeiten), bezogen auf die Gesamtdauer der Äußerung; siehe hierzu Abbildung 5.6 auf Seite 130 bzw. Abbildung 5.9 auf Seite 135 (jeweils untere Bildhälfte). Die aufgetragenen RELATIVEN SILBENANFANGSZEITEN ergeben sich durch Einführung des Divisors der Gesamtdauer bzw. $Tanf_i = 1/(t_{N,3} - t_{1,1}) \cdot t_{i,1}$. Die Silbenanfangszeit der ersten Silbe beträgt $t_{1,1} = 0$, damit vereinfacht sich die Berechnung auf $Tanf_i = 1/t_{N,3} \cdot t_{i,1}$.
Pro Testsatz entsteht aus den N relativen Silbenanfangszeiten ein Vektor $Tanf$ gemäß

folgender Berechnung:

$$Tanf = \begin{pmatrix} Tanf_1 \\ Tanf_2 \\ ..... \\ Tanf_N \end{pmatrix} = \frac{1}{t_{N,3}} \begin{pmatrix} t_{1,1} \\ t_{2,1} \\ .... \\ t_{N,1} \end{pmatrix} = \frac{1}{t_{N,3}} \begin{pmatrix} 0 \\ t_{2,1} \\ .... \\ t_{N,1} \end{pmatrix} \tag{5.9}$$

Die grafische Darstellung gem. Abbildung 5.6 auf Seite 130 bzw. Abbildung 5.9 auf Seite 135 (jeweils untere Bildhälfte) zeigt die gem. Gleichung 5.9 errechneten relativen Silbenanfangszeiten, aufgetragen jeweils über der laufenden Nummer der korrespondierenden Silbe.

Aus Gleichung 5.9 ist sofort ersichtlich, daß die resultierende Flußgrafik stets bei dem Wert 0 beginnt und bei einem Wert kleiner als 100 Prozent endet. Letzteres ergibt sich aus folgender Betrachtung:

$$Tanf_N = \frac{1}{t_{N,3}} \cdot t_{N,1} < 1 \qquad \Leftrightarrow \qquad t_{N,3} > t_{N,1} \tag{5.10}$$

Die MITTLERE (Netto-)SILBENDAUER berechnet sich über Mittelwertbildung der (im Idealfalle $8N$) Silbendauern sämtlicher Testsätze der betrachteten Zielwortkomplexität (siehe Abbildung 5.8 auf Seite 133) bzw. der (im Idealfalle $24N$) Silbendauern aller Testsätze der Sitzung (siehe Abbildung 5.11 auf Seite 138), je nach gewählter Auswertungsmodalität.

Die zusätzlich angegebene MITTLERE NETTOSILBENRATE als ein Maß für das Sprechtempo des produzierten Satzes ist der Kehrwert der mittleren (Netto)Silbendauer. Sie trägt damit die Dimension 1/sec bzw. Hz.

Die MITTLERE BRUTTOSILBENRATE als ein weiteres Maß für das Sprechtempo ist der Kehrwert der mittleren Bruttosilbendauer. Letztere ergibt sich analog zur dargestellten Berechnungsweise der mittleren (Netto-)Silbendauer durch Mittelwertbildung.

5.3.3. Auditive Analyse segmentaler Defizite

Der Segmentierungseditor (siehe 5.4 auf Seite 124) gestattet dem Untersucher die silbenbezogene Angabe von Attributen entsprechend seiner perzeptiven Wahrnehmung bzw. seines auditiven Befundes. Zur Verfügung stehen hierfür die Silbenattribute

- IT Iteration

5. *Realisierung des Moduls SATZPRODUKTION*

- AD Hinzufügung
- E> Elision
- PE Phonetische Entstellung
- PP Phonematische Paraphasie
- VE Verschmelzung.

Die Attribuierung verfolgt hier 2 Ziele: Zum einen sollen alle 9 bzw. 10 tatsächlich zum Testsatz gehörenden Zielsilben aus der Menge der insgesamt detektierten Silben - ggf. angereichert um eventuell fälschlicherweise detektierte Geräuschartefakte - vom System erkannt werden. Da MODIAS nur eine Silbensegmentierung, aber keine Silbenerkennung durchführt und weiterhin vom Untersucher keine phonetische Transkription der detektierten Silbensegmente abverlangt wird, ist dem System die Position der Zielsilben zunächst unbekannt. Die Zuordnung zu den Zielsilben entsteht erst durch die Logik einer korrekten Attribuierung (siehe ergänzend Seite 125).
Zum anderen soll eine differenzierte Betrachtung der möglichen Ursachen für eine gestörte Satzproduktion erfolgen. Hierzu werden die 3 Parameter

- ITERATIONSQUOTIENT
- ADDITIONSQUOTIENT
- PAUSENQUOTIENT

eingeführt. Sie beschreiben die relativen Anteile der für Iterationen, Additionen bzw. Pausen aufgebrachten Zeit bezogen auf die Gesamtdauer der Äußerung (siehe hierzu auch Seite 131).
Die bisherige Betrachtung gem. 5.3.2 auf Seite 155 hatte zunächst sowohl eine korrekte Satzproduktion als auch eine fehlerfreie Segmentierung vorausgesetzt. Die Silbenmatrix S gem. Gleichung 5.6 hatte genau $N = 9$ bzw. $N = 10$ Elemente enthalten (je nach Testsatz), d.h., sämtliche detektierte Silben waren auch Zielsilben.
Treten Hinzufügungen, Iterationen oder Elisionen in der Satzproduktion auf, ist die Anzahl der Zeilen der Silbenmatrix nicht mehr a priori bekannt.
Die hieraus resultierende Silbenmatrix S besitze in der weiteren Betrachtung nunmehr

M Zeilen, wobei M (die Anzahl der Silben bzw. Zeilen) größer, gleich oder aber auch kleiner als die fixe Anzahl N der Zielsilben sein kann.

$$S = \begin{pmatrix} t_{1,1} & t_{1,2} & t_{1,3} \\ t_{2,1} & t_{2,2} & t_{2,3} \\ \cdots & \cdots & \cdots \\ t_{M,1} & t_{M,2} & t_{M,3} \end{pmatrix} = \begin{pmatrix} 0 & t_{1,2} & t_{1,3} \\ t_{2,1} & t_{2,2} & t_{2,3} \\ \cdots & \cdots & \cdots \\ t_{M,1} & t_{M,2} & t_{M,3} \end{pmatrix} \tag{5.11}$$

Der Vektor $Tnetto$ gem. Gleichung 5.7 wird daher ebenfalls um die Nettosilbendauern der überflüssigen bzw. fehlenden Silben auf die Dimension M erweitert, damit gilt auch

$$Tnetto = \begin{pmatrix} Tnetto_1 \\ Tnetto_2 \\ \cdots \\ Tnetto_M \end{pmatrix} = \begin{pmatrix} t_{1,3} - t_{1,1} \\ t_{2,3} - t_{2,1} \\ \cdots \\ t_{M,3} - t_{M,1} \end{pmatrix} = \begin{pmatrix} t_{1,3} \\ t_{2,3} - t_{2,1} \\ \cdots \\ t_{M,3} - t_{M,1} \end{pmatrix} \tag{5.12}$$

Die Gesamtdauer aller vorkommenden Pausen T_{paus} berechnet sich aus den zwischen den Silben $1 \ldots M$ liegenden zeitlichen Lücken oder anders ausgedrückt aus der Differenz zwischen Gesamtdauer der Äußerung $T_{ges} = t_{M,3} - t_{1,1}$ und der Summe aller Nettosilbendauern $Tnetto_i$.

Somit kann T_{paus} wie folgt angegeben werden:

$$T_{paus} = T_{ges} - \sum_{i=1}^{M} Tnetto_i = (t_{M,3} - t_{1,1}) - \sum_{i=1}^{M} Tnetto_i = t_{M,3} - \sum_{i=1}^{M} Tnetto_i \tag{5.13}$$

Der gesuchte PAUSENQUOTIENT Q_{paus} berechnet sich daher zu

$$Q_{paus} = \frac{T_{paus}}{T_{ges}} = \frac{t_{M,3} - \sum_{i=1}^{M} Tnetto_i}{t_{M,3}} = 1 - \frac{\sum_{i=1}^{M} Tnetto_i}{t_{M,3}} \tag{5.14}$$

Die gesamte Dauer der innerhalb des Satzes für Iterationen aufgebrachten Zeit T_{iter} wird aus der Summe der korrespondierenden Nettosilbendauern gewonnen. Zur mathematischen Beschreibung wird der Vektor P eingeführt, der mit seinen m Elementen P_i jeweils die Position aller m mit dem Attribut Iteration versehenen Silben innerhalb des Vektors $Tnetto$ angibt. Damit gilt:

$$T_{iter} = \sum_{i=1}^{m} Tnetto_{P_i} \tag{5.15}$$

Der gesuchte ITERATIONSQUOTIENT errechnet sich somit zu:

$$Q_{iter} = \frac{T_{iter}}{T_{ges}} = \frac{\sum_{i=1}^{m} Tnetto_{P_i}}{t_{M,3}} \tag{5.16}$$

Die gesamte Dauer der innerhalb des Satzes für Hinzufügungen aufgebrachten Zeit T_{add} wird ebenfalls aus der Summe der korrespondierenden Nettosilbendauern gewonnen. Der Vektor P ordnet mit seinen n Elementen P_i jeweils die Position aller n hinzugefügten Silben innerhalb des Vektors *Tnetto* zu. Damit gilt:

$$T_{add} = \sum_{i=1}^{m} Tnetto_{P_i} \tag{5.17}$$

Der gesuchte ADDITIONSQUOTIENT errechnet sich somit zu:

$$Q_{add} = \frac{T_{add}}{T_{ges}} = \frac{\sum_{i=1}^{m} Tnetto_{P_i}}{t_{M,3}} \tag{5.18}$$

5.4. Anwendungen und Ergebnisse

In der Entwicklungsgruppe klinische Neuropsychologie München-Bogenhausen (EKN) wurden innerhalb eines Forschungsprojekts zwei verschiedene Sprechapraxie-Therapieverfahren bezüglich ihrer Wirksamkeit verglichen. Zum Zeitpunkt der Erstellung dieser Arbeit lagen die Ergebnisse von 2 Langzeit-Studien vor, die hierzu bereits eine deutliche Tendenz aufzeigen und daher an dieser Stelle vorgestellt werden.

5.4.1. Untersuchungsziel

Ziel der Untersuchung war es, die Wirksamkeit eines neuen Therapieverfahrens (SYNC) gegenüber einem konventionellen Verfahren (CONV) nachzuweisen. Die neue Methode beruht auf einer Synchronisationsaufgabe aus dem System TUS (z.B. in [Ahr94]), das als *Rhythmische Stimulation* bezeichnet wird.

Für die Untersuchung wurde ein Design gewählt, bei dem die Patienten abwechselnd

mit beiden Methoden therapiert wurden (SYNC-CONV-SYNC-CONV bzw. CONV-SYNC-CONV-SYNC). Vor Therapiebeginn wurde jeweils eine Baseline-Untersuchung mit MODIAS durchgeführt; nach jeder Therapiephase (bestehend aus 8 Untersuchungssitzungen verteilt über 2 Wochen) erfolgte jeweils eine Kontrolluntersuchung.

5.4.2. Stichprobe

Patient AE (männlich, 50 Jahre alt): Schweres Sprechapraxie-Syndrom infolge eines Schlaganfalls (Infarkt der linken mittleren Hirnarterie 9 Monate vor Therapiebeginn).

Patient EM (männlich, 63 Jahre alt): Mittelschweres Sprechapraxie-Syndrom infolge eines Schlaganfalls (Infarkt der linken mittleren Hirnarterie 3 Monate vor Therapiebeginn).

5.4.3. Ergebnisse

Die Ergebnisse sind in Abbildung 5.20 bis 5.23 dargestellt. Bedingt durch die Variation der Reihenfolge der Therapieverfahren entstehen dabei in allen Grafiken zwei Kurvenverläufe, die zueinander versetzt angeordnet sind. Die beiden Therapiephasen sind durch horizontale Balken gekennzeichnet. Die Sternchen markieren in jeder Grafik diejenigen Phasen, in denen sich signifikante Verbesserungen einstellten.

In Abbildung 5.20 sind die gemessenen Gesamtsatzdauern über die 24 Testsätze des Moduls gemittelt. Man sieht, daß beide Patienten die Satzproduktionsaufgabe mit zunehmender Therapiedauer in kürzerer Zeit absolvieren. Der Patient EM zeigt in der letzten Therapiephase (CONV) wieder eine Verlangsamung, allerdings bei weniger segmentalen Fehlern.

In Abbildung 5.21 sind die mittleren Anzahlen segmentaler Fehler pro Satz dargestellt. Auch hier konnte ein kontinuierlicher Rückgang festgestellt werden. Die Grafik zeigt, daß die erste Therapiephase für Patient EM (SYNC) besonders wirksam war; die anschließenden Phasen zeigen nur noch geringfügige Verbesserungen („Deckeneffekt").

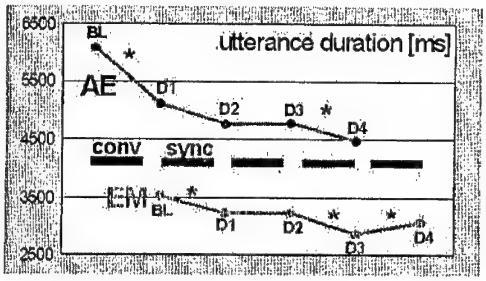

Abbildung 5.20.: Gemessene mittlere Gesamtdauern für die 24 Testsätze des Moduls SATZPRODUKTION. Beide Patienten konnten ihre Testsätze nach erfolgter Therapie jeweils schneller realisieren.

Abbildung 5.22 stellt die mit MODIAS gemessenen Pausenquotienten und Abbildung 5.23 die gemessenen Additionsquotienten dar. Durch Vergleich der beiden Grafiken wird ersichtlich, daß die Verbesserungen in den Satzdauern zum größten Teil auf eine Reduktion der Pausen zurückzuführen sind. Additionen wie z.B. Fehlversuche und Iterationen traten bei beiden Patienten selten auf, so daß keine signifikanten Veränderungen festgestellt werden konnten.

Zusammenfassend kann folgendes festgestellt werden:

1. Beide Therapieverfahren führen zu signifikanten Verbesserungen, d.h., beide Therapieverfahren sind wirksam.
2. Die Überlegenheit der SYNC-Therapie kann sich erst bei einer größeren Stichprobe zeigen. Allerdings ist schon nach den beiden abgeschlossenen Therapie-Studien erkennbar, daß die neue Therapie (SYNC) häufiger zu signifikanten Verbesserungen führt als die herkömmliche Methode.

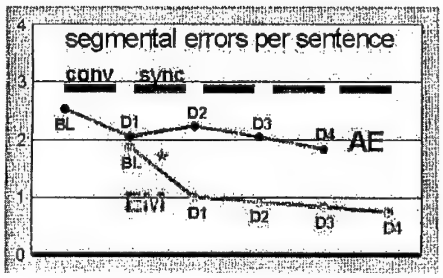

Abbildung 5.21.: Gemessene mittlere Anzahlen segmentaler Fehler pro Testsatz. Beide Patienten konnten ihre Testsätze nach erfolgter Therapie mit weniger segmentalen Fehlern realisieren. Die erste SYNC-Therapie ist besonders wirksam; danach sind nur noch geringfügige Verbesserungen feststellbar.

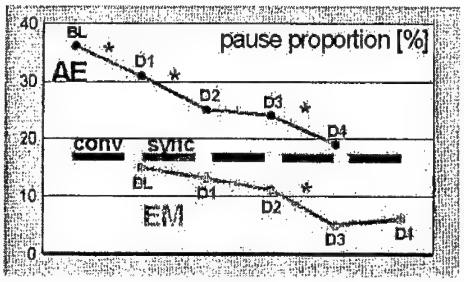

Abbildung 5.22.: Gemessene mittlere Pausenquotienten für die 24 Testsätze. Der Anteil der Sprechpausen an der insgesamt benötigten Realisierungszeit geht im Laufe der Therapie deutlich zurück. Durch Vergleich mit Abbildung 5.23 wird sofort ersichtlich, daß dieser Effekt als Hauptursache für die Verbesserungen in den Satzdauern gesehen werden kann.

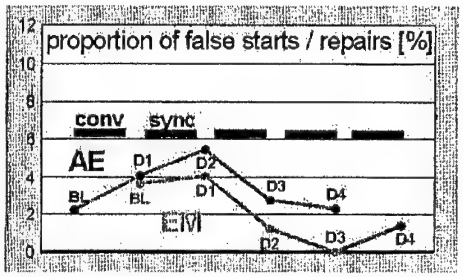

Abbildung 5.23.: Gemessene mittlere Additionsquotienten für die 24 Testsätze. Additionen wie Fehlversuche und Iterationen traten ingesamt sehr selten auf, daher sind keine signifikanten Veränderungen zu erwarten.

6. Realisierung des Moduls SCHNELLE SILBENWIEDERHOLUNGEN

Das diagnostische und technische Konzept zu dem Modul SCHNELLE SILBENWIEDERHOLUNGEN wurde bereits im Kapitel 3 dargestellt. Nach einer kurzen Erläuterung der relevanten diagnostischen Fragestellungen widmet sich dieses Kapitel nun der konkreten Realisierung innerhalb des MODIAS-Systems. Der interdisziplinäre Charakter der Aufgabenstellung läßt hierbei eine Zweiteilung sinnvoll erscheinen in die

- klinisch/diagnostisch relevanten Anteile, wie z.B. Untersuchungsprotokoll, gewähltes Untersuchungsmaterial, Ablauf einer Untersuchungssitzung und Auswertung bzw. Darstellung der Ergebnisse, sowie in
- technisch relevante Anteile, wie z.B. angewandte Verfahren und Algorithmen, die zu der ingenieurwissenschaftlichen Lösung der medizinischen Problemstellung geführt haben.

Der letzte Abschnitt zeigt eine klinische Anwendung des Moduls SCHNELLE SILBENWIEDERHOLUNGEN im Rahmen eines Syndromvergleiches zwischen Dysarthrie und Sprechapraxie. Die Untersuchung wurde in der Entwicklungsgruppe Klinische Neuropsychologie des Städtischen Krankenhauses München-Bogenhausen durchgeführt.

6.1. Diagnostische Fragestellungen

Die Überprüfung schneller Wechselbewegungen (Diadochokinese) stellt einen Standard in der neurologischen Diagnostik motorischer Störungen dar. Diadochokineseleistungen

der Hand liefern ein sensitives Maß für Störungen von motorischer Koordination und Bewegungsgeschwindigkeit ([Her99]). Dysdiadochokinese zählt insbesondere zu den Zeichen ataktischer Bewegungsstörungen ([Not94]).
In der Diagnostik zentraler Sprechstörungen werden rasche Wechselbewegungen durch Silbenwiederholungsaufgaben geprüft. Auch diese Aufgaben bilden ein sensitives Testverfahren für das Vorliegen motorischer Störungen ([Por82]), wenn auch Diadochokineseleistung und sprechmotorisches Defizit dissoziieren können ([Zie96]). Trotz der Bedeutung, die die schnelle Silbenwiederholung in der Diagnostik zentraler Sprechstörungen besitzt, gibt es noch keine apparativen Verfahren zur klinischen Messung artikulatorischer Diadochokineseleistungen. Das in der Praxis übliche time-by-count-Verfahren ist ungenau und erfaßt nur Tempoaspekte, nicht die Regularität der Wiederholungen. Daher wurde dieser Aufgabentyp in das Diagnostiksystem MODIAS integriert ([Merk99a]).

6.2. Aufbau und Gestaltung

6.2.1. Untersuchungsprotokoll und -material

Der Patient wird gebeten, die folgenden Silben möglichst rasch und regelmäßig zu wiederholen:

- */ba/*, */da/*, */ga/*, */na/* sowie die
- */bada/* und */dana/*.

Damit entstehen insgesamt 6 Aufzeichnungen pro Untersuchungssitzung. Die Aufgaben */ba/*, */da/*, */ga/* und */na/* erfordern eine repetitive Verschlußbildung durch die primären Artikulatoren Lippen, Zungenspitze und Zungenrücken, im Falle von */na/* bei zusätzlich abgesenktem Gaumensegel. Die Aufgabe */bada/* erfordert ein repetitives Wechseln zwischen zwei unterschiedlichen Artikulatoren, */dana/* erfordert eine Koordination von Zungen- und Velumbewegungen. Auf die Prüfung der Sequenz */badaga/* wurde verzichtet, weil diese Aufgabenstellung häufig mit fehlerhaften Realisierungen verbunden ist.
Die Untersuchung liefert somit ein Profil, das die Diadochokineseleistung unterschiedlicher Artikulatoren und den Einfluß zusätzlicher Anforderungen an die Bewegungskoordination abbildet.

6.2.2. Untersuchungsablauf und Bedienung

Nach dem Aufruf des MODIAS-Systems erscheint die Startoberfläche gem. Abbildung 4.1 auf Seite 47 mit dem entsprechenden Startbutton für das Modul SCHNELLE SILBENWIEDERHOLUNGEN.
Nach Betätigung des Buttons START! gelangt man unmittelbar zur Organisationsoberfläche gem. Abbildung 4.2 auf Seite 48 mit der Aufforderung zur Eingabe sämtlicher relevanter Daten zum Patienten und zur aktuellen Untersuchungssitzung.
Der Untersucher wird jetzt zur Aufzeichnungsoberfläche gem. Abbildung 6.1 weitergeleitet.

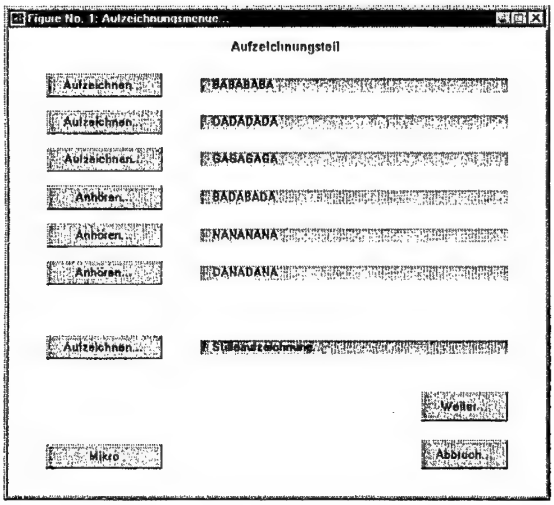

Abbildung 6.1.: Aufzeichnungsoberfläche des Moduls SCHNELLE SILBENWIEDERHOLUNGEN. Die Aufzeichnung wird jeweils durch den Button AUFZEICHNEN... gestartet. Liegt bereits eine Aufzeichnungssitzung vor, ist nur die Wiedergabe über den vom System ersetzten Button ANHÖREN... möglich.

Die Aufzeichnung jeder Silbenwiederholungsaufgabe wird durch Betätigung des entsprechenden Buttons AUFNEHMEN... gestartet. Jetzt wird vom Patienten jede Sil-

benwiederholungsaufgabe mit der ihm maximal möglichen Geschwindigkeit gesprochen. Soll die Aufzeichnung gespeichert werden, genügt das Schließen des Soundrecorders (siehe Abbildung 4.5 auf Seite 51) und die Speicherung wie vorgeschlagen. Der Dateiname ist durch MODIAS bereits voreingestellt, ebenso die Systemeinstellungen bezüglich Samplingrate (22050 Hz) und Auflösung (16Bit).

Die im Untersuchungsprotokoll vorgesehene Stilleaufzeichnung sieht eine Aufzeichnung ohne Aufsprache bzw. ohne jegliche Art von (nicht systemimmanenten) Störgeräuschen mit einer Dauer von ca. 5 sec vor. Eine automatisch durchgeführte statistische Analyse des systemspezifischen Rauschens des Aufzeichnungssystems Soundkarte/Rechner ermöglicht eine wesentlich exaktere Festlegung der Silbengrenzen bzw. von Signalabschnitten mit artikulatorischer Aktivität.

Wie bereits dargestellt, besteht jede Aufzeichnung aus einer Kette von Silbenwiederholungen, die im Regelfall kontinuierlich aneinandergereiht sind, bei Patienten mit Artikulationsstörungen aber durch Sprechpausen unterschiedlicher Länge unterbrochen sein können. Zur Erfassung der suprasegmentalen bzw. zeitlichen Aspekte der artikulatorischen Diadochokinese ist es zunächst erforderlich, die Aufzeichnung weitgehend automatisch in einzelne Silben zu segmentieren. Da die hierzu bekannten Verfahren bei unscharfer Artikulation an ihre Grenzen stoßen, muß neben einem gesteigerten Aufwand in das technische Verfahren dem Untersucher auch die Möglichkeit gegeben werden, in die Segmentierung steuernd eingreifen zu können. Gerade in der zeitlichen Lage der Einzelsilben und in ihrer Dauer liegt die Information, die zur späteren Beurteilung der artikulatorischen Diadochokinese herangezogen werden soll.

Hierzu wurde in dem Modul SCHNELLE SILBENWIEDERHOLUNGEN ein spezieller Segmentierungseditor implementiert, der einen mit der subjektiven Lautheitsempfindung des menschlichen Gehörs korrespondierenden Kurvenverlauf (Lautheitskontur) darstellt und damit eine geeignete Oberfläche für die weitere Bearbeitung durch den Untersucher zur Verfügung stellt.

Die erforderlichen Berechnungen der Lautheitskonturen sowie der daraus abgeleiteten Silbengrenzen werden automatisch beim Übergang von der Aufzeichnungsoberfläche zur anschließenden Segmentierungsoberfläche gem. Abbildung 6.2 für jede der realisierten Silbenwiederholungsaufgaben berechnet.

Die Segmentierungsoberfläche gestattet für jede durchgeführte Aufzeichnung (erkennbar an einer gelben Markierung am rechten Rand der jeweiligen Textfelder) über den Button SEGMENTIEREN den Aufruf des Segmentierungseditors gem. Abbildung 6.3.

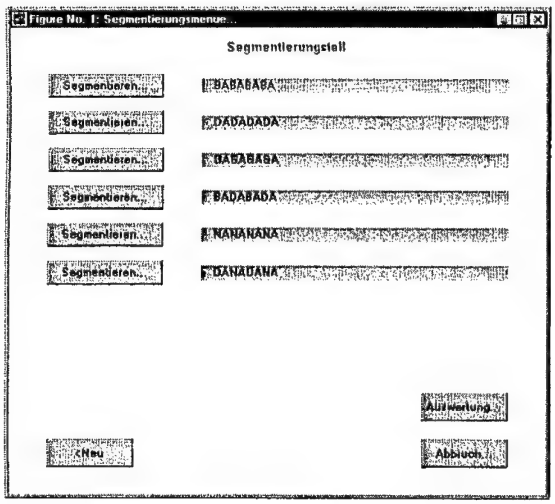

Abbildung 6.2.: Segmentierungsoberfläche des Moduls SCHNELLE SILBENWIEDERHOLUNGEN. Durch Betätigung des Buttons SEGMENTIEREN... wird die Segmentierung der jeweils zugeordneten Silbenwiederholungsaufgabe eingeleitet.

Im Segmentierungseditor werden zunächst die berechneten Lautheitskonturen für die jeweilige Silbenwiederholungsaufgabe grafisch darstellt und dabei die vom Segmentierungsalgorithmus automatisch erkannten Silben entsprechend markiert. Die Lautheitskonturen zeigen idealtypisch lokale Maxima an den Orten der Silbenkerne bzw. lokale Minima an den Orten der Silbenübergänge bzw. in den Sprechpausen.
Das rechts oben eingeblendete Informationsfeld zeigt dem Untersucher den aktuellen Stand der Silbensegmentierung für die gerade bearbeitete Silbenwiederholungsaufgabe an. Die eingeblendete Zahl entspricht der Anzahl der produzierten Silben nach dem gegenwärtigen Kenntnisstand des Systems.

Müssen die vorgeschlagenen Silbengrenzen bezüglich ihrer zeitlichen Lage korrigiert werden, stehen dem Untersucher mehrere Interaktionsmöglichkeiten zur Verfügung. Unterstützend wird hierzu im unteren Teil des Bildschirms das zeitsynchrone Oszillogramm

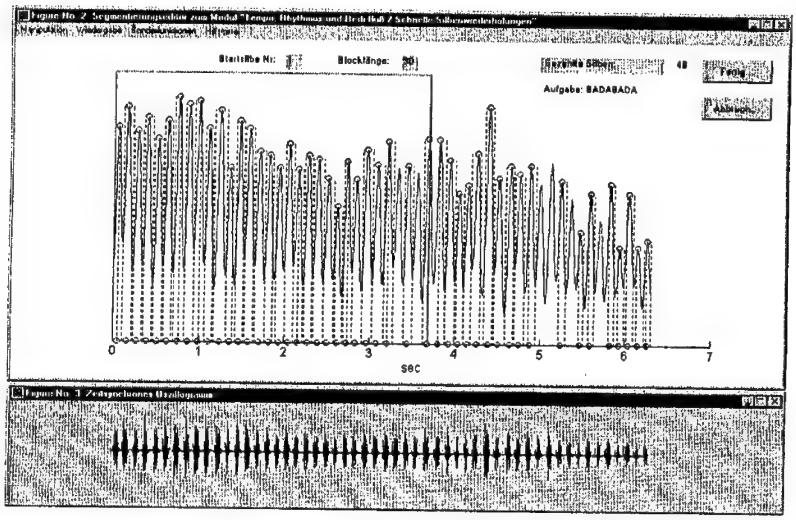

Abbildung 6.3.: Bedienoberfläche des Segmentierungseditors des Moduls SCHNELLE SILBENWIEDERHOLUNGEN. Hier erfolgt die manuelle Einflußnahme des Untersuchers auf die vom System vorgeschlagene Segmentierung. Die projizierte Kurve stellt die subjektive Lautheitsempfindung des menschlichen Gehörs dar, sie eignet sich daher besonders als Grundlage zur Detektion von Silben. Das zeitsynchrone Oszillogramm dient zur Unterstützung bei der manuellen Segmentierung.

gem. Abbildung 6.3 als zusätzliche Visualisierung des Sprachsignales eingeblendet.

Das Menü MANIPULATION dient der unmittelbaren, manuellen Einflußnahme auf die bisher automatisch erfolgte Segmentierung. Es bietet die folgenden Auswahl:

- ZWEI SILBEN VEREINIGEN: Diese Funktion ist immer dann sinnvoll, wenn eine Silbe z.B. aufgrund eines Lautheitseinbruches im Silbenkern vom Segmentierungsalgorithmus fälschlicherweise in zwei (Halb-)Silben geteilt wurde. Hierzu Silbenkern mit linker Mousetaste selektieren, damit wird die aktuelle Silbe mit der unmittelbar nachfolgenden Silbe vereinigt.

- SILBE LÖSCHEN: Gestattet die Entfernung von Silben, z.B. bei fehlinterpretierten Geräuschartefakten. Der betreffende Silbenkern wird mit linker Mousetaste selektiert, damit wird die gesamte Silbe entfernt.
- SILBE SETZEN: Manuelle Markierung einer Silbe, die ggf. vom Algorithmus zunächst unerkannt geblieben war. Hierzu Silbenkern mit linker Mousetaste setzen, die zugehörigen Silbengrenzen werden automatisch vom System vorgeschlagen.
- LINKE bzw. RECHTE SILBENGRENZE verschieben: Erlaubt die manuelle Verschiebung von Silbengrenzen. Der Silbenkern wird hierfür mit der linken Mousetaste selektiert, danach wird der neue Ort der linken bzw. rechten Silbengrenze mit gleicher Mousetaste vorgeben.

Das Menü WIEDERGABE gestattet dem Untersucher die perzeptive Kontrolle der bisherigen Silbensegmentierung. Die Auswahl erlaubt die Wiedergabe der gesamten Aufzeichnung, die manuelle Einzelsilbenwiedergabe (wiederzugebenden Silbe jeweils mit linker Mousetaste selektieren) oder die selektive Wiedergabe derjenigen Silben, die innerhalb des definierten Auswerteblocks liegen (Erläuterung zur Definition von Auswerteblöcken folgt im weiteren Text).

Das Menü SONDERFUNKTIONEN bietet dem Nutzer die Möglichkeit, sogenannte Auswerteblöcke zu definieren. Diese Auswerteblöcke legen fest, welche Silben der Aufzeichnung in die weiteren Auswertungen eingehen bzw. welche nicht mehr weiter betrachtet werden sollen. Diese Option gewährleistet, daß bei der Auswertung eine konstante Anzahl von Silben berücksichtigt wird. Das Auswertungsfenster kann so positioniert werden, daß

- Sprechpausen (z.B. durch Luftholen),
- Sequenzfehler,
- Artefakte (z.B. durch husten, räuspern, lachen usw.)

nicht in die Auswertung eingehen. Nach Aufruf der entsprechenden Auswahl erscheinen am oberen Bildrand des Segmentierungseditors zwei Eingabefelder mit folgender Information:

- STARTSILBE: Laufende Nummer der Silbe, die den Anfang des Auswerteblocks markiert,

- BLOCKLÄNGE: Anzahl der Silben, die innerhalb des Auswerteblocks liegen sollen.

Wird mindestens eines dieser Felder mit der Taste <Return> verlassen, ist die gewählte Blockdefinition aktiviert, zusätzlich wird eine farbige Markierung in die Segmentierungsoberfläche zur grafischen Visualisierung projiziert (siehe Abbildung 6.3).
Die Blockdefinition kann jederzeit geändert oder aber auch über die entsprechende Menüauswahl wieder verworfen werden. Sie führt nach dem Verlassen des Segmentierungseditors nicht zur Beschneidung der Aufzeichnung, d.h., die Anteile außerhalb des gewählten Blockes bleiben erhalten und stehen damit auch beim Wiederaufruf des Segmentierungseditors uneingeschränkt zur Verfügung.

Die Auswahl MANUELLE THRESHOLD-FESTLEGUNG gestattet dem Untersucher die spezifische Vorgabe einer Mindestlautheit für die Detektion von Silben direkt in der Grafik über die linke Mousetaste.

Das Menü HILFSMITTEL ermöglicht eine benutzerdefinierte Zoomfunktion bezüglich der Zeitachse bei fester Skalierung der Lautheitsachse. Die Festlegung des darzustellenden Bereiches erfolgt durch Aufziehen eines Zoomfensters bei gedrückter linker Mousetaste. Die drei weiteren Optionen dieses Menüs steuern das Ein- bzw. Ausblenden des Oszillogrammes sowie die Anpassung der Oszillogramm-Darstellung auf bereits erfolgte Manipulationen (Linke bzw. rechte Silbengrenze verschoben) oder geänderte Skalierungen (Zeitachse gezoomt).
Das letzte Menü ? bietet eine Hilfefunktion mit ausführlicher Beschreibung der Bedienung des Segmentierungseditors.

Nach Verlassen des Segmentierungseditors mit dem Button OK werden die jetzt überflüssigen Signalabschnitte vor der ersten und nach der letzten Silbe entfernt (unabhängig von einem eventuell festgelegten Auswerteblock). Sind alle Aufzeichnungen innerhalb der Segmentierungsoberfläche bearbeitet, erfolgt der Übergang zur Auswertung.

6.2.3. Auswertung, Ergebnisdarstellung und Datenexport

6.2.3.1. Auswertung und grafische Ergebnispräsentation

Der Auswertungsteil des Moduls SCHNELLE SILBENWIEDERHOLUNGEN widmet sich primär der grafischen Visualisierung von Tempo, Rhythmus und Silbenfluß, hier speziell der Bewertung der Fähigkeit zur

- Produktion einer ausreichend hohen Wiederholungsrate,
- Konstanthaltung des Zeitmusters, und zur
- Konstanthaltung des Amplitudenmusters

bei schnellen Silbenwiederholungen. Das erste Auswertungsblatt gem. Abbildung 6.4 stellt hierzu die relevanten Meßgrößen dar. Dabei wurde für beide Grafiken eine Projektion gewählt, die auf der Abszisse ein äquidistantes Raster mit der absoluten Anzahl der berücksichtigten Silben enthält. Waren zuvor für einzelne Silbenwiederholungsaufgaben bei der entsprechenden Vorverarbeitung im Segmentierungseditor Auswerteblöcke definiert worden, so werden hier nur die Silben innerhalb dieser festgelegten Blöcke angezeigt und in alle weiteren Berechnungen miteinbezogen.
Im oberen Teil des Auswertungsblattes sind für jede der 6 Silbenwiederholungsaufgabe die ermittelten BRUTTOSILBENDAUERN in der Dimension ms über der entsprechenden Silbe aufgetragen, also ein Parameter zur Bewertung der zeitlichen Aspekte der Silbenproduktion. Die Bruttosilbendauern ergeben sich aus dem Zeitabstand jeweils zweier aufeinanderfolgender linker Silbengrenzen bzw. aus der Summe von Nettosilbendauern (Zeitabstand jeweils zwischen linker und rechter Silbengrenze) und den sich gegebenenfalls unmittelbar anschließenden Pausen. Der resultierende Verlauf jeder Silbenwiederholungsaufgabe ist hier in einer eigenen Farbe dargestellt gemäß der unten rechts dargestellten Legende.
Das hier gezeigte Beispiel gem. Abbildung 6.4 entstammt einer Untersuchung an einem sprechgesunden Probanden. Der Silbenfluß ist kontinuierlich ohne relevante Auffälligkeiten bei einer insgesamt hohen Realisierungsgeschwindigkeit.
Die eingeblendete gestrichelte Linie zeigt den Durchschnittswert der Bruttosilbendauern an, berechnet über alle realisierten Silben (Grand Average), hier 164 ms.

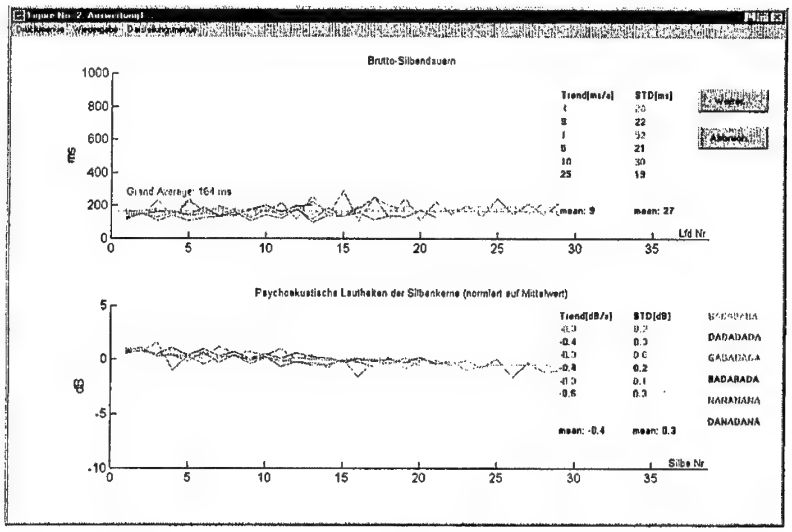

Abbildung 6.4.: Auswertung zum Modul SCHNELLE SILBENWIEDERHOLUNGEN, Blatt 1. Die Kurven geben Aufschluß über die zeitlichen Aspekte der Silbenproduktion. Sämtliche Verläufe werden statistisch bezüglich linearer Trends und ihrer mittleren Schwankung analysiert.

Die am rechten oberen Rand dargestellte Tabelle gibt an, ob die dargestellten Kurvenverläufe jeweils einen (linearen) Trend aufweisen. Diese aus der Zeitreihe der Silbendauern ermittelten Trends im Verlauf der Bruttosilbendauern werden in der Dimension ms/sec angegeben, wobei ein positives Vorzeichen einen Anstieg der Bruttosilbendauern (Tempoverlangsamung) anzeigt, ein negatives Vorzeichen hingegen einen Abfall der Bruttosilbendauern (Tempoerhöhung).
Die in der rechten Spalte der Tabelle aufgelisteten Werte geben zusätzlich an, wie stark die dargestellten Kurvenverläufe zeitlich schwanken. Sie ergeben sich durch Berechnung der (trendbereinigten) Standardabweichung der gemessenen Bruttosilbendauern und sind in der Dimension ms angegeben. Ein hoher Wert steht hier für eine hohe Schwankung der Bruttosilbendauern (geringe Regularität), ein niedriger Wert entsprechend für eine geringe Schwankung (hohe Regularität). Für den Sprechgesunden ist wie im

gezeigten Beispiel eine hohe Regularität zu erwarten, also glatte Kurvenverläufe und damit auch vergleichsweise geringe Standardabweichungen.
Im unteren Teil des Auswertungsblattes sind die normierten PSYCHOAKUSTISCHEN LAUTHEITEN in der Dimension dB über der jeweils korrespondierenden Silbe aufgetragen, also ein Parameter zur Bewertung der phonatorischen Aspekte der Silbenproduktion. Sie sind normiert auf die mittlere Lautheit aller Silben der jeweiligen Silbenwiederholungsaufgabe. Auch hier ist der zu jeder Aufgabe gehörenden Kurvenverlauf farblich gemäß der unten rechts dargestellten Legende eindeutig zuordnenbar.
Die am rechten unteren Bildrand eingeblendete Tabelle gibt analog zum oberen Teil des Auswertungsblattes für die Verläufe der normierten psychoakustischen Silbenlautheiten die gegebenenfalls unterliegenden Trends in der Dimension dB/sec an, wobei hier ein positives Vorzeichen mit einer Zunahme der Silbenlautheiten bei fortschreitender Realisierungsdauer korrespondiert. Ein negatives Vorzeichen deutet hingegen auf eine Abnahme der Silbenlautheiten hin, die z.B. einen Hinweis auf eine unökonomische Exspiration geben kann.
Die in der rechten Spalte der Tabelle aufgeführten Zahlenwerte errechnen sich aus den (trendbereinigten) Standardabweichungen der psychoakustischen Lautheiten sämtlicher Silben der jeweiligen Silbenwiederholungsaufgabe. Ein hoher Wert steht hier für hohe Schwankungen der phonatorischen Intensität (niedrige Stabilität), ein niedriger Wert entsprechend für eine geringe Schwankung (hohe Stabilität).
Das Menü DRUCKEN stellt diverse Varianten der Druckausgabe zur Verfügung wie Farbdruck, Schwarz-Weiß-Druck oder Kopie in die Zwischenablage für die weitere Verwendung in klinischen Berichten usw.
Das Menü WIEDERGABE erlaubt dem Untersucher die akustische Wiedergabe einzelner Silbenwiederholungsaufgaben bei gleichzeitiger Betrachtung der hierzu ermittelten Kurvenverläufe.
Das Menü DARSTELLUNG gestattet die selektive Projektion einzelner, gerade näher zu betrachtender Silbenwiederholungsaufgaben. Bei Detailauswertungen kann diese optionale Beeinflussung der Darstellung eine verbesserte Übersichtlichkeit herbeiführen.

Das zweite Auswertungsblatt gem. Abbildung 6.5 liefert einen Scatterplot bzw. die Anzeige von Datenpunkten in einer zweidimensionalen Ebene, die jeweils 1 Parameterpaar pro Einzelsilbe repräsentieren. Damit ist die Beobachtung von Parameterkorrelationen sehr anschaulich möglich.

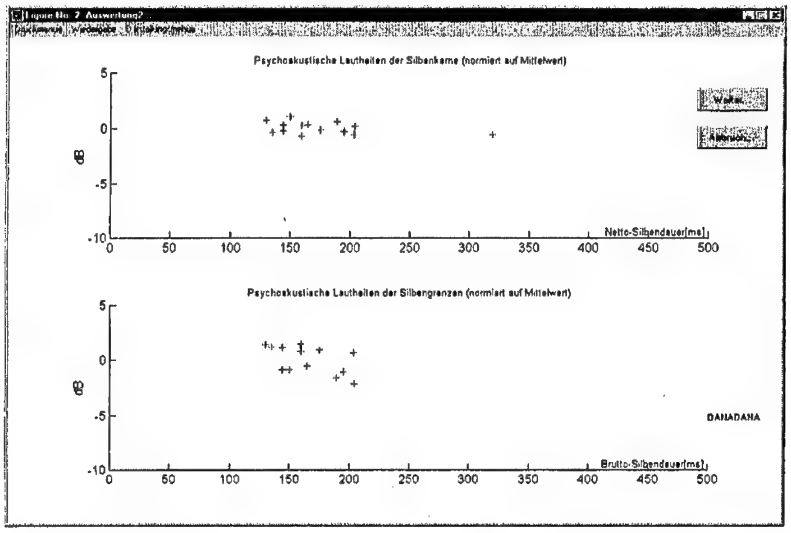

Abbildung 6.5.: Auswertung zum Modul SCHNELLE SILBENWIEDERHOLUNGEN, Blatt 2. Die dargestellten Scatterplots erlauben dem Untersucher die Beobachtung von Korrelationen zwischen jeweils 2 akustischen Parametern.

In der oberen Bildhälfte wird hierzu pro Silbe ein Datenpaar projiziert, welches sich aus der jeweiligen

- Nettosilbendauer in der Dimension ms und der
- normierten psychoakustischen Lautheit des Silbenkerns in der Dimension dB

zusammensetzt. Je höher die Konstanz der Nettosilbendauern und der zugehörigen Silbenkernlautheiten wäre, desto stärker würde sich hier die Konzentration der Datenpunkte in der Darstellungsebene ausprägen. Der Schwerpunkt liegt hierbei also auf der Beurteilung von *Unregelmäßigkeiten in der Silbenproduktion.*

In der unteren Bildhälfte wird für jede realisierte Silbe ebenfalls ein Datenpaar in eine zweidimensionale Ebene projiziert, hier jedoch gebildet aus der jeweiligen

- Bruttosilbendauer in der Dimension ms und der
- normierten psychoakustischen Lautheit der jeweils linken Silbengrenze in der Dimension dB.

Diese Darstellung bietet zum einen den Vorteil, daß durch die Projektion der Bruttosilbendauer auch die eventuell vorhandenen Pausen zwischen den einzelnen Silben mit eingehen. Damit wird der Fall einer zwar konstanten Nettosilbendauer, aber einer variablen Pausendauer sichtbar. Der Schwerpunkt liegt hier also auf der Beurteilung von *Unregelmäßigkeiten im Silbenfluß.*
Der andere Vorteil dieser Darstellung ist die spezifischere Beurteilung der komplexen Silbenfolge */dana/* mit ihrer höheren artikulatorischen Anforderung. Bei der Silbenfolge */dana/* ist bei korrekter Gaumensegelfunktion und damit vollständiger Verschlußbildung des Rachenraumes gegenüber dem Nasenraum beim Plosivlaut */d/* ein Unterschied in der psychoakustischen Lautheit gegenüber */a/* zu erwarten. Während bei der Bildung des Konsonanten */d/* sowohl der Mundraum als auch der Nasenraum abgeschlossen sind und das Sprachsignal daher für kurze Zeit den Wert 0 annimmt, wirkt bei der Bildung des Nasals */n/* durch die Absenkung des Gaumensegels der Nasenraum als Resonator und die abgestrahlte Signalenergie hat einen deutlich von 0 verschiedenen Wert. In Abbildung 6.5 ist dieser Effekt durch eine Aufteilung der Datenpunkte in zwei Gruppen unterschiedlicher Lautheit erkennbar.
Das Menü DRUCKEN stellt diverse Varianten der Druckausgabe zur Verfügung wie Farbdruck, Schwarz-Weiß-Druck oder Kopie in die Zwischenablage für die weitere Verwendung in klinischen Berichten usw.
Das Menü WIEDERGABE erlaubt dem Untersucher die akustische Wiedergabe einzelner Silbenwiederholungsaufgaben bei gleichzeitiger Betrachtung der korrespondierenden Scatterplots.

Das Menü DARSTELLUNG gestattet zwei verschiedene Darstellungsvarianten:

- Beschränkung der Scatterplots auf einzelne, ggf. näher zu betrachtende Silbenwiederholungsaufgaben,

- Beschränkung der Darstellung auf Mittelwerte und Standardabweichungen, berechnet über alle realisierten Silben der jeweiligen Silbenwiederholungsaufgaben anstelle der Standardeinstellung der Einzelsilbendarstellung.

Beide Varianten sind bei Detailauswertungen hilfreich und verschaffen dem Untersucher eine bessere Übersicht.
Das dritte und letzte Auswertungsblatt des Moduls SCHNELLE SILBENWIEDERHOLUNGEN zeigt die Abbildung 6.6. Es bietet dem Untersucher einen Gesamtüberblick über Tempo und Redefluß der Silbenproduktion.

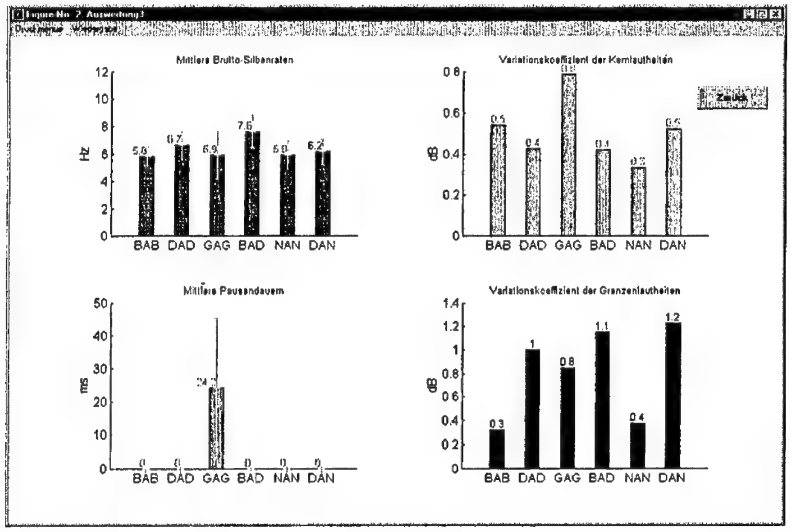

Abbildung 6.6.: Auswertung zum Modul SCHNELLE SILBENWIEDERHOLUNGEN, Blatt 3. Die angezeigten Parameter fassen die bisherigen Einzelergebnisse zusammen und bieten dem Untersucher so einen schnellen Überblick bezüglich der artikulatorischen Diadochokinesefähigkeit des Patienten.

Hierzu werden 4 verschiedene akustische Parameter getrennt nach Silbenwiederholungsaufgaben abgeleitet und dargestellt, das sind die

- Mittleren Bruttosilbenraten: Kehrwerte der mittleren Bruttosilbendauern in der Dimension Hz, somit das korrespondierende Geschwindigkeitsmaß der Silbenproduktion;

- Mittlere Pausendauer: Durchschnittliche Pausendauer zwischen den einzelnen Silben in der Dimension ms. Zugleich mittlere Differenz zwischen Brutto- und Nettosilbendauern;

- Variationskoeffizienten der Kernlautheiten: Standardabweichung der normierten psychoakustischen Lautheiten der Silbenkerne in der Dimension dB;

- Variationskoeffizienten der Grenzenlautheiten: Standardabweichung der normierten psychoakustischen Lautheiten der jeweils linken Silbengrenzen in der Dimension dB.

Die beiden Variationskoeffizienten entstehen durch Relativierung der entsprechenden Standardabweichungen am Mittelwert, um eventuelle Abhängigkeiten zwischen Mittelwert und Streuung zu eliminieren.
Typisch für einen Sprechgesunden ist hier wie im Beispiel gem. Abbildung 6.6 die höchste gemessene Silbenrate bei der Silbenwiederholungsaufgabe */bada/* sowie die größte gemesse Variation der Silbengrenzenlautheit bei */dana/*. Bei einem sprechgestörten Patienten wird in der Praxis bei den Silbenraten häufig eine ähnliche Verteilung bezüglich der einzelnen Aufgaben wie bei einem Sprechgesunden beobachtet, jedoch insgesamt auf einem niedrigeren Niveau.

Das Menü DRUCKEN stellt diverse Varianten der Druckausgabe zur Verfügung wie Farbdruck, Schwarz-Weiß-Druck oder Kopie in die Zwischenablage für die weitere Verwendung in klinischen Berichten usw.
Das Menü WIEDERGABE erlaubt dem Untersucher die akustische Wiedergabe bei gleichzeitiger Betrachtung der Auswertegrafiken.

6.2.3.2. Automatischer Datenexport

Mit Aufruf der Auswertung und damit synchron zur Darstellung des ersten Auswertungsblattes gem. Abbildung 6.4 werden automatisch zwei Datenfiles mit allen relevanten Ergebnissen im entsprechenden Arbeitsverzeichnis angelegt:

- Silbenfile: Datenfile mit der Endung *...sil.txt*; es enthält im wesentlichen Silbendauern und Lautheiten jeder realisierten Einzelsilbe der 6 Silbenwiederholungsaufgaben;
- Statistikfile: Datenfile mit der Endung *...sta.txt*; es enthält im wesentlichen die Mittelwerte der Silbendauern und Lautheiten, berechnet für jede der Silbenwiederholungsaufgaben. Wurden Auswerteblöcke (siehe 6.2.2) definiert, gehen nur die Silben innerhalb dieser Blöcke in die statistischen Berechnungen ein.

Beide Datenfiles sind ASCII-codiert und damit universell zur weiteren Nutzung oder Weiterverarbeitung mit alternativen Softwareprodukten einsetzbar.

Format des Silbenfiles: Das Silbenfile enthält ein Zahlenfeld in Matrixdarstellung, wobei jeder Silbenwiederholungsaufgabe eine separate Matrixzeile zugeordnet ist. Jede Zeile enthält in den ersten vier Feldern (Spalten) folgende Informationen:

- Feld 01: Stets mit der Zahl 1 belegt
- Feld 02: Silbenwiederholungsaufgabe, codiert gem. Tabelle 6.1
- Feld 03: Gesamtdauer der Silbenwiederholungsaufgabe in ms (vom Beginn der ersten bis zum Ende der letzten Silbe der Aufzeichnung)
- Feld 04: Anzahl der segmentierten Silben innerhalb der Aufzeichnung

Pro realisierter Silbe der laufenden Nummer i wird nun jeweils ein 3-elementiger Daten-Record angehängt. Damit ergeben sich für die Felder Nr. $4i+1 \ldots 4i+4$ die folgenden Belegungen:

- Feld $4i + 1$: (Absolute) Startzeit der Silbe Nr. i in ms
- Feld $4i + 2$: (Netto-)Silbendauer der Silbe Nr. i in ms
- Feld $4i + 3$: Psychoakustische Lautheit an der linken Grenze der Silbe Nr. i in dB
- Feld $4i + 4$: Psychoakustische Lautheit im Kern der Silbe Nr. i in dB

Das Silbenfile bzw. die enthaltene Datenmatrix verfügt also über eine variable Anzahl von Spalten, die durch die Silbenwiederholungsaufgabe mit der größten Anzahl realisierter bzw. segmentierter Silben bestimmt wird.

Format des Statistikfiles: Das Statistikfile enthält ebenfalls ein Zahlenfeld in Matrixdarstellung, wobei hier nicht mehr jede Silbe einzeln, sondern nur noch abgeleitete statistische Größen als repräsentative Parameter für die Gesamtheit aller Silben der Silbenwiederholungsaufgabe eingehen. Die Statistik umfaßt nur diejenigen Silben, die innerhalb der eingestellten Auswerteblöcke liegen (siehe hierzu 6.2.2 auf Seite 171). Wurde kein Auswerteblock gewählt, werden alle Silben der Aufzeichnung berücksichtigt.
Jede Zeile dieser Matrix enthält in den ersten vier Feldern (Spalten) in Analogie zur vorhergehenden Darstellung die folgenden Informationen:

- Feld 01: Stets mit der Zahl 1 belegt
- Feld 02: Silbenwiederholungsaufgabe, codiert gem. Tabelle 6.1
- Feld 03: Gesamtdauer des eingestellten Auswerteblocks in ms (vom Beginn der ersten bis zum Ende der letzten Silbe des Auswerteblocks)
- Feld 04: Anzahl der segmentierten Silben innerhalb des Auswerteblocks

Die weiteren Felder (bzw. Spalten der Matrix) sind mit folgenden Parametern belegt (gültig jeweils nur für die Silben innerhalb der vordefinierten Auswerteblöcke):

- Feld 05: Linearer Trend der Bruttosilbendauern in ms/sec (positives Vorzeichen bei Zunahme der Bruttosilbendauern)

Silbenwiederholungsaufgabe	Code
/ba/	1
/da/	2
/ga/	3
/bada/	4
/na/	5
/dana/	6

Tabelle 6.1.: Identifizierung der Silbenwiederholungsaufgabe anhand der Codierung im Feld 02 von Silben- und Statistikfile

- Feld 06: Standardabweichung der trendbereinigten Bruttosilbendauern in ms
- Feld 07: Mittlere Bruttosilbenrate (entsteht durch Mittelung der Kehrwerte der Bruttosilbendauern) in Hz
- Feld 08: Standardabweichung der Bruttosilbenrate in Hz
- Feld 09: Mittlere Pausendauer in ms
- Feld 10: Standardabweichung der Pausendauer in ms
- Feld 11: Linearer Trend der psychoakustischen Lautheiten in den Silbenkernen
- Feld 12: Standardabweichung der trendbereinigten psychoakustischen Lautheiten in den Silbenkernen
- Feld 13: Standardabweichung der psychoakustischen Lautheiten in den Silbenkernen
- Feld 14: Standardabweichung der psychoakustischen Lautheiten an den linken Silbengrenzen

Das Statistikfile bzw. die darin enthaltene Datenmatrix verfügt also im Gegensatz zum Silbenfile über eine feste Anzahl von (14) Spalten.

6.3. Technische Realisierung: Entwickelte Verfahren und Algorithmen

Die zu lösende technische Aufgabenstellung liegt in dem Modul SCHNELLE SILBEN-WIEDERHOLUNGEN in der akustischen Analyse der artikulatorischen Diadochokinese, also der Fähigkeit, eine vorgegebene Silbenfolge so rasch wie möglich zu wiederholen. Dabei liegt der Schwerpunkt in der

- akustischen Analyse der Wiederholungsrate,
- akustischen Analyse der Regelmäßigkeit des Zeitmusters,
- akustischen Analyse der Kontinuität der Amplituden in den Silbenkernen und an den Silbengrenzen.

Unter dem Begriff *Zeitmuster* soll im Folgenden die Information über die zeitliche Lage von Silbenkernen, Silbengrenzen und Silbendauern verstanden werden. Der Lösungsansatz zur Analyse des Zeitmusters führt wie schon im vorangehenden Kapitel in das Gebiet der Spracherkennung, da hier unabhängig vom gewählten Erkennungsprozeß bereits in der Vorverarbeitungsstufe die Problematik der Halbsilben-, Silben- oder Einzelwortsegmentierung betrachtet werden muß ([Lea80]).
Die Erkennung von Silbengrenzen streift zusätzlich eine typische Problematik der Biosignalverarbeitung, nämlich die korrekte Detektion von Nutzsignalen in verrauschter Signalumgebung. Von Interesse ist hier die Beantwortung der Frage, wann sich meßbare akustische Signalparameter deterministisch und signifikant ändern bei einer gleichzeitig möglichst hohen Robustheit gegenüber stochastischen Signalveränderungen z.B. durch Rauschpeaks, Artefakte usw.

6.3.1. Detektion von Silbenkernen, Silbengrenzen und Sprechpausen

Der vorgestellte Segmentierungsalgorithmus fußt auf einer von RUSKE und BEHAM in [Rus92] vorgestellten Vorverarbeitung durch Berechnung der modifizierten Lautheitsfunktion. Durch Anwendung dieses Verfahrens entstehen Parameter-Zeitverläufe, die die

menschliche Lautheitsempfindung über eine Modellierung der physiologischen Vorgänge im Gehör nachbilden und damit für die nachfolgende algorithmische Silbensegmentierung eine optimale Detektionsgrundlage bereitstellen.
Das Berechnungsverfahren zur Gewinnung der beiden Lautheitsverläufe und zur Ableitung der Orte von Silbenkernen und Silbengrenzen ist identisch zu Kapitel 5.3.1 auf Seite 142 ff. und wird daher hier nicht mehr weitergehend erläutert. Die erforderlichen psychoakustischen Grundlagen des Hörens, die Beschreibung des Funktionsmodells der Lautheitsempfindung mit der Ableitung der Lautheitsfunktion und der modifizierten Lautheitsfunktion sowie die optimale Detektion von Silbenkernen und Silbengrenzen in verrauschter Signalumgebung finden auch in diesem Modul Anwendung.

6.3.2. Analyse des Zeitrasters schneller Silbenwiederholungen

Nach Abschluß der automatischen - ggf. auch manuell korrigierten - Segmentierung sind die Orte der Silbenkerne und der Silbengrenzen festgelegt. Das MODIAS-System ist nun in der Lage, akustische Ereignisse mit silbischem Charakter innerhalb der Aufzeichnung zu detektieren und sowohl in Bezug auf die Ereigniszeit als auch in Bezug auf die relevante psychoakustische Empfindungsgröße der Lautheit zu analysieren.
Für die folgende Betrachtung liege eine akustische Repräsentation einer realisierten Silbenwiederholungsaufgabe vor. Die Segmentierung in K Silben sei korrekt erfolgt. Die weitere Darstellung zeigt die schrittweise Ermittlung der akustischen Parameter

- Nettosilbendauer,
- Bruttosilbendauer,
- Normierte psychoakustische Lautheit im Silbenkern,
- Normierte psychoakustische Lautheit an der linken Silbengrenze

auf, die für die Ergebnispräsentation in variierender Kombination Verwendung finden und eine Grundlage für weitere Analysen bilden.
MODIAS erstellt für jede Silbenwiederholungsaufgabe eine Silbenmatrix S, deren Zeilen mit den Silben i korrespondieren. Die 3 Spalten beinhalten die detektierten Zeitpunkte für Silbenbeginn ($t_{i,1}$), Silbenkern ($t_{i,2}$) und Silbenende ($t_{i,3}$). Per definitionem beginnt

jede Aufzeichnung mit der ersten Silbe bei $t = 0$, d.h., für alle weiteren Betrachtungen gilt $t_{1,1} = 0$.

$$S = \begin{pmatrix} t_{1,1} & t_{1,2} & t_{1,3} \\ t_{2,1} & t_{2,2} & t_{2,3} \\ \cdots & \cdots & \cdots \\ t_{K,1} & t_{K,2} & t_{K,3} \end{pmatrix} = \begin{pmatrix} 0 & t_{1,2} & t_{1,3} \\ t_{2,1} & t_{2,2} & t_{2,3} \\ \cdots & \cdots & \cdots \\ t_{K,1} & t_{K,2} & t_{K,3} \end{pmatrix} \tag{6.1}$$

Die in der Auswertung des MODIAS-Modules u.a. dargestellte NETTOSILBENDAUER ergibt sich durch silbenweise Differenzbildung zwischen den Zeitpunkten von Silbenende und Silbenanfang bzw. $Tnetto_i = t_{i,3} - t_{i,1}$.
Damit entsteht pro Silbenwiederholungsaufgabe aus den K Nettosilbendauern ein Vektor $Tnetto$ gemäß folgender Berechnungsweise:

$$Tnetto = \begin{pmatrix} Tnetto_1 \\ Tnetto_2 \\ \cdots \\ Tnetto_K \end{pmatrix} = \begin{pmatrix} t_{1,3} - t_{1,1} \\ t_{2,3} - t_{2,1} \\ \cdots \\ t_{K,3} - t_{K,1} \end{pmatrix} = \begin{pmatrix} t_{1,3} \\ t_{2,3} - t_{2,1} \\ \cdots \\ t_{K,3} - t_{K,1} \end{pmatrix} \tag{6.2}$$

Der Parameter BRUTTOSILBENDAUER ergibt sich durch Differenzbildung jeweils zwischen den Zeitpunkten aufeinanderfolgender Silbenbeginne bzw. $Tbrutto_i = t_{i+1,1} - t_{i,1}$.
Damit entsteht pro Silbenwiederholungsaufgabe aus den $K - 1$ Bruttosilbendauern ein Vektor $Tbrutto$:

$$Tbrutto = \begin{pmatrix} Tbrutto_1 \\ Tbrutto_2 \\ \cdots \\ Tbrutto_{K-1} \end{pmatrix} = \begin{pmatrix} t_{2,1} - t_{1,1} \\ t_{3,1} - t_{2,1} \\ \cdots \\ t_{K,1} - t_{K-1,1} \end{pmatrix} = \begin{pmatrix} t_{2,1} \\ t_{3,1} - t_{2,1} \\ \cdots \\ t_{K,1} - t_{K-1,1} \end{pmatrix} \tag{6.3}$$

Die psychoakustischen Lautheiten an den Orten der Silbenkerne und der Silbengrenzen ergeben sich beide aus dem Verlauf der Lautheit $N(t)$. Hierbei ist anzumerken, daß bei der zuvor durchgeführten Segmentierung die Orte der Silbenkerne durch die Detektion lokaler Maxima der modifizierten Lautheitsfunktion $N_m(t)$ bestimmt worden waren, die Orte der Silbengrenzen jedoch durch Detektion lokaler Minima der korrespondierenden Lautheitsfunktion $N(t)$.
Durch Einsetzen der Ereigniszeitpunkte der Silbenkerne entsteht pro Silbenwiederholungsaufgabe der Vektor $Nkern$, bestehend aus den K Lautheiten der Silbenkerne

$N(t_{1,2}) \dots N(t_{K,2})$ durch folgenden Zusammenhang:

$$Nkern = \begin{pmatrix} Nkern_1 \\ Nkern_2 \\ \cdots\cdots\cdots \\ Nkern_K \end{pmatrix} = \begin{pmatrix} N(t_{1,2}) \\ N(t_{2,2}) \\ \cdots\cdots\cdots \\ N(t_{K,2}) \end{pmatrix} \tag{6.4}$$

Nach Normierung und Logarithmierung ergibt sich schließlich der gesuchte akustische Parameter der NORMIERTEN PSYCHOAKUSTISCHEN LAUTHEITEN an den Orten der Silbenkerne $\dot{N}kern$ (dB-skaliert):

$$\dot{N}kern = 10\log\left\{\frac{1}{\frac{1}{K}\sum_{i=1}^{K} Nkern_i}\begin{pmatrix} Nkern_1 \\ Nkern_2 \\ \cdots\cdots\cdots \\ Nkern_K \end{pmatrix}\right\} = 10\log\left\{\frac{1}{\frac{1}{K}\sum_{i=1}^{K} Nkern_i}\begin{pmatrix} N(t_{1,2}) \\ N(t_{2,2}) \\ \cdots\cdots\cdots \\ N(t_{K,2}) \end{pmatrix}\right\} \tag{6.5}$$

Die psychoakustischen Lautheiten an den Orten der jeweils linken Silbengrenzen ergeben sich analog durch Einsetzen der Ereigniszeitpunkte der Silbengrenzen in die Lautheitsfunktion $N(t)$. Damit entsteht pro Silbenwiederholungsaufgabe der Vektor $Ngren$, bestehend aus den K Lautheiten an den linken Silbengrenzen $N(t_{1,1}) \dots N(t_{K,1})$ durch folgenden Zusammenhang:

$$Ngren = \begin{pmatrix} Ngren_1 \\ Ngren_2 \\ \cdots\cdots\cdots \\ Ngren_K \end{pmatrix} = \begin{pmatrix} N(t_{1,1}) \\ N(t_{2,1}) \\ \cdots\cdots\cdots \\ N(t_{K,1}) \end{pmatrix} \tag{6.6}$$

Nach Normierung und Logarithmierung ergibt sich schließlich der gesuchte akustische Parameter der normierten psychoakustischen Lautheiten an den Orten der linken Silbengrenzen $\dot{N}gren$ (dB-skaliert):

$$\dot{N}gren = 10\log\left\{\frac{1}{\frac{1}{K}\sum_{i=1}^{K} Ngren_i}\begin{pmatrix} Ngren_1 \\ Ngren_2 \\ \cdots\cdots\cdots \\ Ngren_K \end{pmatrix}\right\} = 10\log\left\{\frac{1}{\frac{1}{K}\sum_{i=1}^{K} Ngren_i}\begin{pmatrix} N(t_{1,1}) \\ N(t_{2,1}) \\ \cdots\cdots\cdots \\ N(t_{K,1}) \end{pmatrix}\right\} \tag{6.7}$$

Mit diesen 4 abgeleiteten Parametern lassen sich die beiden ersten Auswertungsblätter des Moduls gem. Abbildung 6.4 und 6.5 hinreichend interpretieren. Zusätzlich werden

in Abbildung 6.4 noch die beiden Kurvenverläufe für Bruttosilbendauer und Silbenkernlautheit jeweils einer statistischen Verlaufsanalyse unterzogen, um sowohl eine Aussage über einen eventuellen Auf- bzw. Abwärtstrend (Trendparameter *Trend*) bzw. über die mittlere Schwankung um diesen Trend (Schwankungsparameter *STD*) treffen zu können. Der Trendparameter ergibt sich durch eine lineare Regressionsanalyse der entsprechenden Verläufe, er trägt die Dimension ms/sec bzw. dB/sec.
Der Schwankungsparameter ergibt sich durch Berechnung der Standardabweichung der jeweils trendbereinigten Kurven. Er trägt damit die Dimension ms bzw. dB.

Der in Abbildung 6.6 grafisch dargestellte Parameter MITTLERE BRUTTOSILBENRATE $mRbrutto$ in der Dimension Hz ist der Kehrwert der mittleren Bruttosilbendauer in Sekunden und stellt damit ein Maß für das Sprechtempo dar. Letztere ergibt sich ihrerseits durch Mittelwertbildung über alle $i = 1 \ldots K-1$ Bruttosilbendauern $Tbrutto_i$ der jeweiligen Silbenwiederholungsaufgabe.

$$mRbrutto = \frac{1}{\frac{1}{K}\sum_{i=1}^{K} Tbrutto_i} \tag{6.8}$$

Der Parameter MITTLERE PAUSENDAUER $mTpaus$ spiegelt die durchschnittliche Pausendauer zwischen den einzelnen Silben wider. Die Pausendauern zwischen zwei jeweils aufeinanderfolgender Silben lassen sich durch Differenzbildung zwischen den Zeitpunkten vom jeweiligen Ende der vorhergehenden zum Anfang der darauffolgenden Silben bestimmen, bzw. gilt $Tpaus_i = t_{i+1,1} - t_{i,3}$.
Damit entsteht pro Silbenwiederholungsaufgabe aus den $K-1$ Pausendauern ein Vektor $Tpaus$:

$$Tpaus = \begin{pmatrix} Tpaus_1 \\ Tpaus_2 \\ \ldots\ldots\ldots \\ Tpaus_{K-1} \end{pmatrix} = \begin{pmatrix} t_{2,1} - t_{1,3} \\ t_{3,1} - t_{2,3} \\ \ldots\ldots\ldots\ldots \\ t_{K,1} - t_{K-1,3} \end{pmatrix} \tag{6.9}$$

Für die gesuchte mittlere Pausendauer $mTpaus$ gilt dann:

$$mTpaus = \frac{1}{K}\sum_{i=1}^{K} Tpaus_i \tag{6.10}$$

Der Parameter VARIATIONSKOEFFIZIENT DER KERNLAUTHEITEN $vNkern$

stellt ein Schwankungsmaß für den zeitlichen Verlauf der normierten und logarithmierten psychoakustischen Lautheiten der Silbenkerne $\dot{N}kern_i$ dar. Der Variationskoeffizient ist in statistischem Sinne ein Streuungsmaß, welches die Standardabweichung am Mittelwert relativiert ([Bor99]). Da die Lautheiten $\dot{N}kern_i$ aufgrund der Berechnung gem. Gleichung 6.5 bereits mittelwertfrei sind, vereinfacht sich die Berechnung von $v\dot{N}kern$ zu:

$$v\dot{N}kern = \sqrt{\frac{1}{K}\sum_{i=1}^{K}\dot{N}kern_i^2} \tag{6.11}$$

Der Parameter VARIATIONSKOEFFIZIENT DER (linken) KERNLAUTHEITEN $v\dot{N}gren$ stellt analog ein Schwankungsmaß für den zeitlichen Verlauf der normierten und logarithmierten psychoakustischen Lautheiten $\dot{N}gren_i$ der jeweils linken Silbengrenzen dar. Bedingt durch die Mittelwertsfreiheit von $\dot{N}gren_i$ gemäß der Berechnung nach Gleichung 6.7 gilt:

$$v\dot{N}gren = \sqrt{\frac{1}{K}\sum_{i=1}^{K}\dot{N}gren_i^2} \tag{6.12}$$

6.4. Anwendungen und Ergebnisse

In der Entwicklungsgruppe klinische Neuropsychologie des Städtischen Krankenhauses München wurde zunächst in einer Pilotstudie der Alterseffekt der Diadochokineseleistung an sprechgesunden Probanden geprüft. Im Anschluß daran folgte eine zweite Studie mit dem Ziel, die Diadochokineseleistung von Patienten mit neurogenen Sprechstörungen mit der gesunder Probanden zu vergleichen.
Die Ergebnisse der Pilotstudie sind in Abbildung 6.7 dargestellt.
Sie können wie folgt zusammengefaßt werden:

1. Die Silbenwiederholungsrate ist am höchsten im jungen Erwachsenenalter, danach fällt sie mit steigendem Alter kontinuierlich ab.
2. Männer erreichen insgesamt höhere Silbenwiederholungsraten als Frauen (Regressionslinie oberhalb!).
3. Der Alterseffekt ist bei Männern schwächer ausgeprägt (Regressionslinie flacher!).

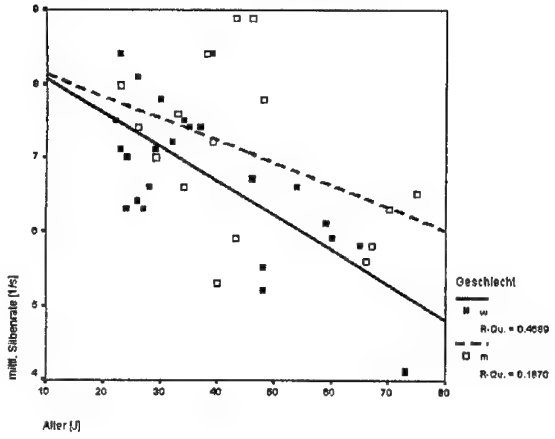

Abbildung 6.7.: Ergebnisse der Pilotstudie zur Prüfung von Alterseffekten bei der artikulatorischen Diadochokinese. Aufgetragen sind das Alter (Abszisse) gegenüber der erzielten Silbenwiederholungsrate (Ordinate). Die Gruppe *weiblich* ist mit ausgefüllten Datenpunkten (durchgezogenen Regressionslinie) markiert, die Gruppe *männlich* entsprechend mit unausgefüllten Datenpunkten (gestrichelte Regressionslinie).

6.4.1. Untersuchungsziel

Ziel der Folgeuntersuchung war es, die Diadochokineseleistung von zwei Patientengruppen mit Dysarthrie- bzw. Sprechapraxiesyndrom zu vergleichen und einer Kontrollgruppe sprechgesunder Probanden gegenüberzustellen.

6.4.2. Stichprobe bzw. Patientengruppe

Es wurden gesunde Kontrollpersonen im Alter zwischen 20 und 75 Jahren sowie Patienten mit Dysarthrie und Sprechapraxie untersucht.

6.4.3. Ergebnisse

Die bisher noch nicht publizierten Ergebnisse der Folgestudie (durchgeführt durch Bettina Brendel und Wolfram Ziegler, EKN,1999) sind in Abbildung 6.8 dargestellt.

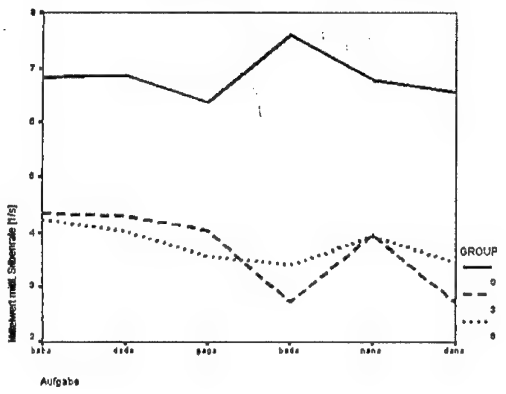

Abbildung 6.8.: Die Abbildung zeigt die Ergebnisse der Folgestudie (Syndromvergleich). Aufgetragen sind die Silbenwiederholungsaufgaben (Abszisse) gegenüber der erzielten Silbenwiederholungsrate (Ordinate). Dargestellt sind die 3 resultierenden Kurvenverläufe für Sprechgesunde (durchgezogen) sowie für Dysarthrie- (gepunktet) und Sprechapraxiepatienten (gestrichelt).

Bei den sprechgesunden Probanden zeigt sich ein Kurvenverlauf, der sehr anschaulich interpretiert werden kann. Die Silbenfolge */ba/* enthält den bilabialen Verschlußlaut */b/*, bei dem beide Lippen verschlossen sind. Rhythmus und Geschwindigkeit werden hier im wesentlichen durch die Trägheit der Lippen und in geringem Umfang auch durch Hebung und Senkung des Unterkiefers beeinflußt. Die Silbenfolge */da/* hingegen enthält den alveolaren Verschlußlaut */d/*, der durch den hochbeweglichen vorderen Zungenteil realisiert wird. Damit werden bei */da/* geringfügig höhere Wiederholungsraten erzielt als bei */ba/*.
Bei der Silbenfolge */ga/* wird der intermittierende Verschluß durch den hinteren Teil der Zunge (Zungenrücken) realisiert. Durch die größere Masse und damit die geringere

Beweglichkeit des Zungenrückens ist die Wiederholungsrate bei */ga/* geringer als bei */ba/* und */da/*.
Die komplexere Silbenfolge */bada/* stellt im Prinzip eine Mischung der Silbenfolgen */ba/* und */da/* dar. Die Bewegungen von Lippen und Zunge müssen hier jedoch genau mit der halben Geschwindigkeit realisiert werden, um eine gleiche Wiederholungsrate zu erzielen. Durch Koartikulation tritt ein Beschleunigungseffekt auf, da während der Lippenöffnung die Zunge schon zum artikulatorischen Zielpunkt wandert. Dadurch ergibt sich bei dieser Aufgabe die höchste Wiederholungsrate.
Die Silbenfolge */na/* erfordert die gleiche Verschlußbildung bzw. Artikulationsstelle wie */da/*, nur ist hier das Gaumensegel ständig abgesenkt und der Nasenraum ist als zusätzlicher Resonator an den Mundraum gekoppelt. Die Wiederholungsrate ist bei */na/* in etwa gleich wie bei */da/*, da das Tempo in beiden Fällen durch die Beweglichkeit der Zungenspitze bestimmt ist.
Bei der Silbenfolge */dana/* ist dieselbe Zungenbewegung wie bei */da/* und */na/* erforderlich. Bei dieser Folge muß jedoch das Gaumensegel im Wechsel für die Silbe */da/* gehoben und für die Silbe */na/* gesenkt werden. Das Gaumensegel bewegt sich daher zeitlich perfekt koordiniert mit einer Frequenz, die der halben Silbenwiederholungsrate entspricht. Offenbar bewirkte diese Koordinationsanforderung bei den gesunden Probanden keine wesentliche Verlangsamung gegenüber den weniger komplexen Aufgaben */da/* und */na/*.

Zusammenfassend bedeutet das für die zu erwartenden Silbenwiederholungsraten (bei sprechgesunden Probanden), daß die Silbenfolgen */da/*, */na/* und */dana/* in etwa gleich schnell, die Silbenfolge */ga/* hingegen langsamer und die Silbenfolge */bada/* schneller realisiert werden. Die mit MODIAS durchgeführten Messungen gem. Abbildung 6.8 bestätigen exakt diese Erwartung.

Bei Patienten mit neurogenen Sprechstörungen tritt ein sehr interessanter Effekt auf. Sofort erkennbar ist die bei allen Silbenwiederholungsaufgaben festgestellte Verlangsamung gegenüber der Normgruppe. Dies gilt für beide Patientengruppen.
Die Patientengruppe der Sprechapraktiker zeigte jedoch einen markanten Profilunterschied. Während bei Sprechgesunden - wie bereits dargestellt - die Silbenwiederholungsrate durch den koartikulativen Effekt bei */bada/* deutlich höhere Werte erreicht hat als bei */ga/*, fällt die Wiederholungsrate bei Sprechapraxiepatienten hingegen deutlich ab.

Der gleiche Effekt tritt bei */dana/* auf.
Patienten mit Sprechapraxiesyndrom bewältigen demnach die Silbenwiederholungsaufgaben mit einfachen Silbenfolgen (mit einem einfachen repetitiven Programm) vergleichsweise leicht. Probleme treten offensichtlich immer dann auf, wenn artikulatorische Wechselbewegungen stattfinden. Dies kann als Beleg für die Vermutung einer Programmierungsstörung als Grundlage des Syndroms der Sprechapraxie gewertet werden ([Zie91]).

7. Realisierung des Moduls FRIKATIVARTIKULATION

Das diagnostische und technische Konzept des Moduls FRIKATIVARTIKULATION wurde bereits im Kapitel 3 dargestellt. Nach einer kurzen Erläuterung der relevanten diagnostischen Fragestellungen widmet sich dieses Kapitel nun der konkreten Realisierung innerhalb des MODIAS-Systems. Der interdisziplinäre Charakter der Aufgabenstellung läßt hierbei eine Zweiteilung sinnvoll erscheinen in die

- klinisch/diagnostisch relevanten Anteile, wie z.B. Untersuchungsprotokoll, gewähltes Untersuchungsmaterial, Ablauf einer Untersuchungssitzung und Auswertung bzw. Darstellung der Ergebnisse sowie in
- technisch relevanten Anteile, wie z.B. entwickelte technische Verfahren und Algorithmen, die zu der ingenieurwissenschaftlichen Lösung der klinischen Problemstellung geführt haben.

Der Weg von der Problemstellung zur technischen Problemlösung führt in diesem Modul über die Verfahren der digitalen Sprachsignalanalyse hinaus in den Bereich der künstlichen Intelligenz. Ausgehend von dem Ansatz, die Güte der Frikativartikulation sowohl auditiv über Expertenurteile als auch akustisch durch Analyse der Sprachsignaleigenschaften beurteilen zu wollen, wird zunächst der Entwurf und die Realisierung von zwei ausgewählten Mustererkennern vorgestellt. Ein Mustererkenner wurde dabei auf die Klassifikation von Frikativen gegenüber ausgewählten Vokalen optimiert, der andere Mustererkenner auf die Klassifikation der einzelnen Frikativtypen untereinander.
Dargestellt werden dabei die typischen Phasen eines Mustererkenner-Designs von

der Entwurfsstrategie über die Vorverarbeitung und Merkmalsextraktion bis hin zur Klassifikationsstrategie. Die eigentliche Beurteilung der Frikativartikulation erfolgt über die Beurteilung des Klassifikationsverhaltens dieser Mustererkenner, also auf der Basis einer weitestgehend automatisierten akustisch-phonetischen Analyse.
Parallel dazu wird ein Verfahren vorgestellt, welches zusätzlich die auditive Prüfung durch Expertenurteile gestattet und einen direkten Vergleich zwischen den beiden Entscheidern, also Mensch und Maschine, ermöglicht.
Jede der beiden Mustererkenner-Realisierungen erforderte entsprechendes Training an sprechgesunden Probanden. Diese systematische und überwachte Phase der Gewinnung von akustischen Referenzmerkmalen (supervised learning) ist für den späteren Erkennungsprozeß von fundamentaler Bedeutung, da die Stichprobe selbst repräsentativ sein muß und die hieraus erhobenen Daten artefaktfrei sein müssen. Die Gewinnung von Trainingsdaten anhand Sprechgesunder wurde im Rahmen einer Reihenuntersuchung an der Universität der Bundeswehr München und der Entwicklungsgruppe Klinische Neuropsychologie München-Bogenhausen durchgeführt.

7.1. Diagnostische Fragestellungen

Der Hauptteil der Diagnostik dysarthrischer Störungen beruht auf einer auditiven Analyse der Äußerungen eines Patienten. Sie läßt sich systematisch in die

- Diagnostik sichtbarer Störungen (z.B. pathologischer Atmungstyp, Störungen von Tonus und Bewegungsfähigkeit von Lippen und Kiefer),
- Diagnostik phonetischer Störungsmerkmale der Funktionskreise Sprechatmung, Phonation und Artikulation,
- Bewertung der Verständlichkeit

untergliedern. Die beiden erstgenannten Varianten erlauben keine Rückschlüsse auf die daraus resultierenden kommunikativen Einschränkungen, während die Bewertung der *Verständlichkeit* wohl eines der wichtigsten Kriterien hierfür darstellt. Die Verständlichkeit vieler Patienten mit dysarthrischen Störungen ist zum Teil erheblich vermindert.

Ein Hauptproblem in der Verständlichkeitsbewertung ist die Vielzahl an Faktoren, die das Verstehen eines dysarthrischen Patienten beeinflussen können, z.B. die Hörfähigkeit und Aufmerksamkeit des Hörers, die Vertrautheit mit dem Patienten, die Vertrautheit mit dem Meßverfahren oder auch die akustische Qualität der Sprachübertragung ([Zie98a]). Die Bewertung ist, auch infolge dieser Faktoren, subjektiv und häufig fehlerbehaftet.
Die maschinelle Spracherkennung bietet prinzipiell die Möglichkeit, die Äußerungen des Patienten durch einen Computer objektiv und mit konstanten Kriterien identifizieren zu lassen. Die bisherigen maschinellen Verfahren sind jedoch nicht ausreichend valide ([Zie98a]).
Das vorliegende MODIAS-Modul ist ein Versuch, anhand sorgfältig gewählter Einzelwörter und beschränkt auf die Lautklasse der Frikative, die Fähigkeit des Patienten zur

- kontrastreichen Artikulation von 3 verschiedenen Frikativen gegenüber einer jeweils vokalischen Lautumgebung (*Diskriminanz Vokal/Frikativ*) und auch zur
- ausreichenden Unterscheidung dieser 3 Frikative (*Frikativ-Differenzierung*)

anhand der Verhaltens spezifischer Mustererkennungssysteme zu bemessen. Aufgrund der ungewissen Validität wurde parallel dazu eine auditive Komponente implementiert, die einen unmittelbaren Vergleich der beiden Entscheidungsinstanzen *Mensch* und *Maschine* gestattet.

7.2. Aufbau und Gestaltung

7.2.1. Untersuchungsprotokoll und -material

Der Patient hat in diesem Modul die Aufgabe, die beiden ausgewählten Einzelwort-Ensembles

- *lassen, Laschen, lachen* bzw.
- *Massen, Maschen, machen*

nachzusprechen. Jedes Einzelwort wird dabei je 3 mal wiederholt. Die beiden Einzelwort-Ensembles sind so zusammengestellt, daß

- die Frikative /s/, /ʃ/ und /x/ jeweils zu gleichen Teilen vorkommen,
- jedem Frikativ der Kontrastvokal /a/ unmittelbar vorangeht,
- jeweils nur der betrachtete Frikativ bei ansonsten fester Lautumgebung variiert.

Der Untersuchungsgang sieht grundsätzlich 2 Durchgänge für die beiden Einzelwort-Ensembles vor; damit entstehen pro Untersuchungssitzung 12 Aufzeichnungen mit insgesamt 36 Einzelwörtern bzw. jeweils 12 Einzelwörtern pro Frikativtyp.

7.2.2. Untersuchungsablauf und Bedienung

Nach dem Aufruf des MODIAS-Systems erscheint die Startoberfläche gem. Abbildung 4.1 auf Seite 47 mit dem entsprechenden Startbutton für das Modul FRIKATIVARTIKULATION.
Nach Betätigung des Buttons START! gelangt man unmittelbar zur Organisationsoberfläche gem. Abbildung 4.2 auf Seite 48 mit der Aufforderung zur Eingabe sämtlicher relevanter Daten zum Patienten und zur aktuellen Untersuchungssitzung.
Der Untersucher wird jetzt zur Steuerungsoberfläche gem. Abbildung 7.1 weitergeleitet, die eine Festlegung bezüglich des weiteren Fortgangs der Untersuchung fordert bzw. ermöglicht.
Im oberen Teil der Steuerungsoberfläche kann der Untersucher auswählen, ob das

- komplette Standardprotokoll mit 2 Durchgängen (12 Aufzeichnungen mit insgesamt 36 Einzelwörtern bzw. je 12 pro Frikativtyp) oder aber das
- verkürzte Untersuchungsprotokoll mit 1 Durchgang (6 Aufzeichnungen mit insgesamt 18 Einzelwörtern bzw. je 6 pro Frikativtyp)

genutzt werden soll.
Im unteren Teil der Steuerungsoberfläche wird der weitere Ablauf bzw. die Zielrichtung der Sitzung festgelegt. Prinzipiell sind hier zwei unabhängige Varianten möglich:

- PATIENTENSITZUNG: Erstellung neuer Aufzeichnungen mit dem zu untersuchenden Patienten und anschließender akustischer Analyse.

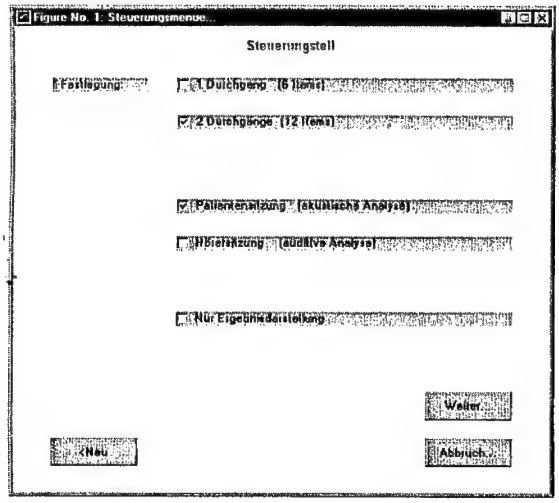

Abbildung 7.1.: Steuerungsoberfläche des Moduls FRIKATIVARTIKULATION. An dieser Stelle wird vom Untersucher der weitere Ablauf der Sitzung und der Umfang des Untersuchungsprotokolls festgelegt bzw. direkt zur Auswertung (bei bereits vorliegenden Ergebnissen) navigiert.

- HÖRERSITZUNG: Befundung bereits vorhandener Aufzeichnungen durch einen geschulten Hörer. Der weitere Ablauf ist in 7.2.2.2 ab Seite 208 beschrieben.

- ERGEBNISDARSTELLUNG: Sichtung ggf. bereits vorhandener Ergebnisse auditiver oder akustischer Analysen. Diese Option ist nur möglich, wenn das entsprechende Textfeld rechts außen gelb markiert ist und somit alle Voraussetzungen vom System geprüft wurden und erfüllt sind. Der weitere Ablauf hierzu ist in 7.2.3 ab Seite 209 erläutert.

Nach Betätigung des Buttons WEITER wird das Untersuchungsprotokoll entsprechend der gewählten Eingabe verzweigt.

7.2.2.1. Patientensitzung (akustische Analyse)

Nach Entscheidung für die Patientensitzung erscheint unmittelbar die erste von bis zu 2 Aufzeichnungsoberflächen gem. Abbildung 7.2.

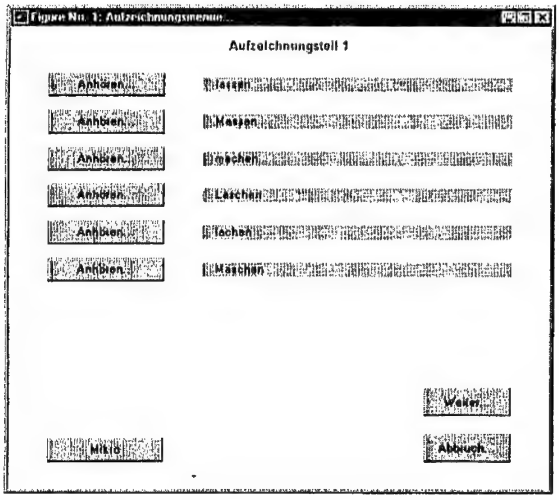

Abbildung 7.2.: Eine der bis zu 2 aufeinanderfolgenden Aufzeichnungsoberflächen des Moduls FRIKATIVARTIKULATION. Die Reihenfolge der Wörter ist randomisiert. Die Aufzeichnung wird jeweils durch den Button AUFZEICHNEN gestartet.

Die Darstellung der Einzelwörter erfolgt randomisiert; die Aufzeichnung wird jeweils durch Betätigung des entsprechenden Buttons AUFNEHMEN initiiert.
Bei der Aufzeichnung wird jedes Einzelwort vom Patienten dreimal hintereinander nachgesprochen mit kurzen, dazwischenliegenden Pausen. Soll die Aufzeichnung gespeichert werden, genügt das Schließen des Soundrecorders (siehe Abbildung 4.5 auf Seite 51) und die Speicherung wie vorgeschlagen. Der Dateiname ist durch MODIAS bereits voreingestellt, ebenso die Systemeinstellungen bezüglich Samplingrate (22050 Hz) und Auflösung (16Bit).
Die im Untersuchungsprotokoll vorgesehene Stilleaufzeichnung sieht eine Aufzeichnung

ohne Aufsprache bzw. ohne jegliche Art von (nicht systemimmanenten) Störgeräuschen mit einer Dauer von ca. 5 sec vor. Aufgrund der akustisch/phonetischen Eigenschaften der Frikative kommt dieser Stilleaufzeichnung in diesem Modul eine entscheidende Bedeutung zu: Die 3 betrachteten Frikative /s/, /ʃ/ und /x/ weisen spektrale Charakteristika auf, die

- nur geringfügige Unterschiede aufweisen,
- im wesentlichen rauschartig sind und
- eine geringe Signalenergie besitzen.

Jede spektrale Färbung des Systemrauschens kann hier schon zu markanten Veränderungen der akustischen Eigenschaften energiearmer Frikative führen. Um hierdurch keine systembedingte Unschärfe der Detektionsleistung nachgeschalteter Mustererkenner entstehen zu lassen, wird anhand der Stilleaufzeichnung eine statistische Analyse des systemspezifischen Rauschens (dominiert vom Aufzeichnungssystem Soundkarte/Rechner) durchgeführt. Die später abgeleiteten akustischen Merkmale der relevanten Laute können so vor dem Erkennungsprozeß in Grenzen korrigiert werden.
Wie bereits dargestellt, besteht jede Aufzeichnung aus drei Einzelwörtern mit dazwischenliegenden Pausenabschnitten. Bedingt durch die ähnlichen akustischen Eigenschaften von (rauschstörungsbehafteten) Pausen und Frikativen müssen für die weitere Verarbeitung die Pausensegmente ohne artikulatorische Aktivität möglichst exakt entfernt werden.
Hierzu wurde in dem Modul FRIKATIVARTIKULATION ein Segmentierungseditor implementiert, der einen mit der subjektiven Lautheitsempfindung des menschlichen Gehörs korrespondierenden Kurvenverlauf (Lautheitskontur) darstellt und damit eine geeignete Oberfläche für die weitere Bearbeitung durch den Untersucher bildet.
Die erforderlichen Berechnungen der Lautheitskonturen sowie der daraus abgeleiteten Wortgrenzen werden automatisch beim Übergang von den Aufzeichnungsoberflächen zur anschließenden Segmentierungsoberfläche gem. Abbildung 7.3 für jede Einzelwortaufzeichnung berechnet.
Die Segmentierungsoberfläche gestattet für jede durchgeführte Aufzeichnung (erkennbar an einer gelben Markierung am rechten Rand der Textfelder) über den Button SEGMENTIEREN den Aufruf des Segmentierungseditors gem. Abbildung 7.4.

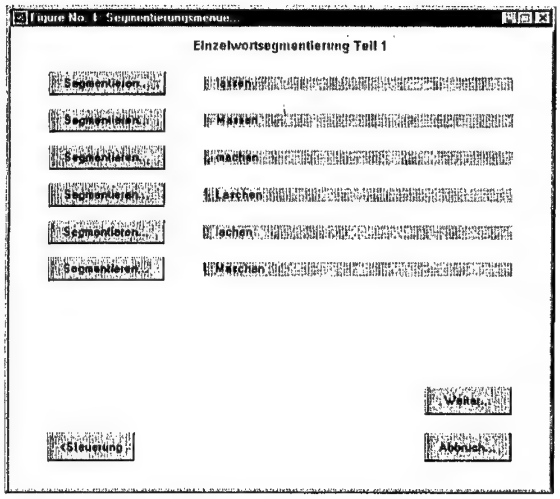

Abbildung 7.3.: Segmentierungsoberfläche des Moduls FRIKATIVARTIKULATION. Durch Betätigung des Buttons SEGMENTIEREN wird die Einzelwortsegmentierung der jeweils zugeordneten Aufzeichnung eingeleitet.

Im Segmentierungseditor wird die für die jeweilige Aufzeichnung errechnete Lautheitskontur grafisch dargestellt, zusätzlich werden die detektierten Grenzen der realisierten Einzelwörter markiert.
Die rechts oben eingeblendeten Informationsfelder zeigen an, wie viele Einzelwörter zum gegenwärtigen Stand der Segmentierung vom System erkannt werden.
Charakteristisch für den Verlauf der Lautheitskontur ist der Einbruch in den Wortmitten, da die stimmlosen Frikative nur einen Bruchteil an Energie gegenüber den umgebenden Vokalen besitzen und damit auch eine wesentlich geringere Lautheit aufweisen. Sind die Frikative bezüglich ihrer artikulatorischen Intensität besonders schwach ausgeprägt, können durch den zu Grunde liegenden Segmentierungsalgorithmus fälschlicherweise auch zwei einzelne Wörter detektiert werden.
Müssen die vorgeschlagenen Segmentgrenzen, wie z.B. in diesem Fall, bezüglich ihrer zeitlichen Lage korrigiert werden, stehen dem Untersucher innerhalb der Oberfläche mehrere Interaktionsmöglichkeiten zur Verfügung. Unterstützend wird ihm hierzu im unteren Teil

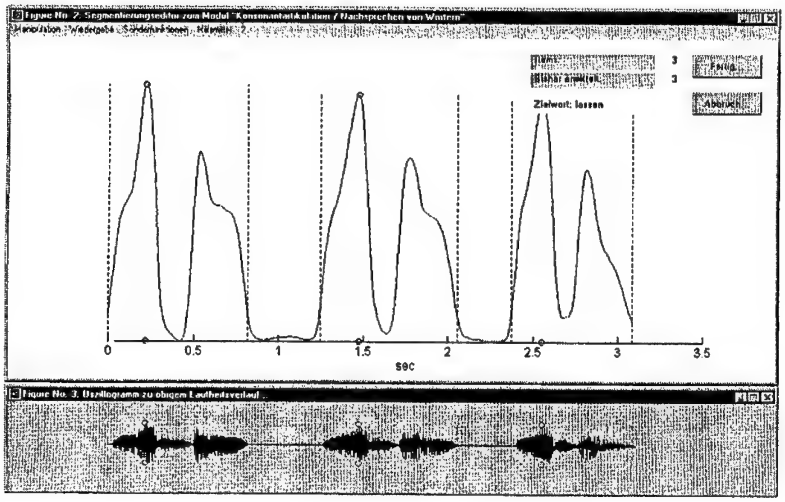

Abbildung 7.4.: Bedienoberfläche des Segmentierungseditors des Moduls FRIKATIVARTIKULATION zur Segmentierung der 3 enthaltenen Einzelwörter. Unterstützend wird hierbei im unteren Teil das zeitsynchrone Oszillogramm eingeblendet.

des Bildschirms eine alternative Darstellung in Form des zeitsynchronen Oszillogramms eingeblendet.
Das Menü MANIPULATION dient der unmittelbaren, manuellen Einflußnahme auf die bisher automatisch erfolgte Segmentierung. Es bietet die folgende Auswahl:

- ZWEI SEGMENTE VEREINIGEN: Diese Funktion ist immer dann sinnvoll, wenn ein Wort wegen des medialen Frikativs einen so starken Lautheitseinbruch aufweist, daß vom Algorithmus fälschlicherweise zwei einzelne Wortsegmente erkannt wurden. Hierzu (den linken) Segmentkern mit linker Maustaste selektieren, danach wird das gewählte Segment automatisch mit dem unmittelbar nachfolgenden zu einem gemeinsamen Segment bzw. Wort vereinigt.
- SEGMENT BZW. WORT LÖSCHEN: Gestattet die Entfernung von Wörtern oder Wortsegmenten, z.B. bei fehlinterpretierten Artefakten. Der betreffende Segment-

kern wird mit linker Mousetaste selektiert, damit wird das gesamte Segment entfernt.

- WORT SETZEN: Manuelle Markierung eines Wortes oder Wortsegmentes, das ggf. vom Algorithmus zunächst unerkannt geblieben war. Hierzu Segmentkern mit linker Mousetaste setzen, die zugehörigen Segmentgrenzen werden automatisch vom System vorgeschlagen.
- LINKE bzw. RECHTE GRENZE VERSCHIEBEN: Erlaubt die manuelle Verschiebung von Wort- bzw. Wortsegmentgrenzen. Der entsprechende Segmentkern wird hierfür mit der linken Mousetaste selektiert, danach wird der neue Ort der linken bzw. rechten Segmentgrenze mit gleicher Mousetaste vorgegeben.

Das Menü WIEDERGABE gestattet dem Untersucher die perzeptive Kontrolle der bisherigen Segmentierung. Die Auswahl erlaubt die Wiedergabe der gesamten Aufzeichnung, die Wiedergabe der detektierten Einzelwörter oder auch die manuelle Einzelwortwiedergabe (wiederzugebenden Segmentkern jeweils mit linker Mousetaste selektieren).

Das Menü SONDERFUNKTIONEN bietet eine Möglichkeit, den Algorithmus zur Berechnung der Lautheitskontur an die jeweilige Sprechgeschwindigkeit des Patienten anzupassen. Werden zunächst weniger als 3 Wörter detektiert und liegt subjektiv ein erhöhtes Sprechtempo vor, kann hier die Anpassung an hohe Sprechgeschwindigkeit erfolgversprechend sein. Werden umgekehrt mehr als 3 Wörter erkannt und liegt subjektiv ein verlangsamtes Sprechtempo vor, sollte eine Anpassung der Sprechgeschwindigkeit nach unten erfolgen. Die Auswahl MANUELLE THRESHOLD-FESTLEGUNG gestattet dem Untersucher die spezifische Vorgabe einer Mindestlautheit für die Detektion von Wortkandidaten direkt in der Grafik über die linke Mousetaste.

Das Menü HILFSMITTEL ermöglicht eine benutzerdefinierte Zoomfunktion bezüglich der Zeitachse bei fester Skalierung der Lautheitsachse. Die Festlegung des darzustellenden Bereiches erfolgt durch Aufziehen eines Zoomfensters bei gedrückter linker Mousetaste. Die drei weiteren Auswahlen dieses Menüs steuern das Ein- bzw. Ausblenden des Oszillogrammes sowie die Anpassung der Oszillogramm-Darstellung auf bereits erfolgte Manipulationen (linke bzw. rechte Segmentgrenze verschoben) oder

geänderte Skalierungen (Zeitachse gezoomt).

Das Menü ? bietet eine Hilfefunktion mit ausführlicher Beschreibung der Bedienung des Segmentierungseditors.
Nach Verlassen des Segmentierungseditors werden die jetzt überflüssigen Signalabschnitte vor dem ersten und nach dem letzten Wort entfernt. Die beiden verbleibenden Pausen zwischen den Einzelwörtern bleiben zwar in der ursprünglichen Aufzeichnung erhalten, gehen aber nicht in die anschließende Bewertung bzw. Auswertung mit ein.
Sind alle Aufzeichnungen segmentiert, wird die Segmentierungsoberfläche mit dem Button WEITER verlassen.

Das Untersuchungsprotokoll geht jetzt von der Phase der Vorverarbeitung in eine Phase der *akustischen Bewertung* über. Unter diesem Begriff soll im folgenden eine Bearbeitung des Sprachsignals verstanden werden, die aufgrund ihres Erkenntnisgewinns für den Durchführenden weder eindeutig der Vorverarbeitung noch der Auswertung zuzuordnen ist. Sie bereitet notwendigerweise die spätere Auswertung beider Aspekte, der Frikativ-Differenzierung und der Diskriminierung Frikativ/Vokal, vor und liefert überdies zum letztgenannten Aspekt detaillierte Informationen.
Der folgende Bildschirm zeigt die erste von 3 Bewertungsoberflächen gem. Abbildung 7.5.
Sie fassen die während der Aufzeichnung randomisiert dargebotenen Einzelwortaufzeichnungen nach den enthaltenen Frikativen zusammen und gestatten über den Button BEWERTUNG die Initiierung der jeweiligen akustischen Bewertung in einer eigenen Bedienoberfläche, dem *grafischen Bewertungseditor* gem. Abbildung 7.6.

Innerhalb des Bewertungseditors wird dem Untersucher ein Kurvenverlauf projiziert, der an dieser Stelle genauer erläutert werden muß. Im Hintergrund, bzw. nicht erkennbar für den Betrachter, arbeiten speziell für die Aufgabe der optimalen Trennung der beiden Lautklassen *Frikativ/Vokal*, also der 3 vorkommenden Klassenpaare

- /s/ ⇔ /a/
- /ʃ/ ⇔ /a/
- /x/ ⇔ /a/

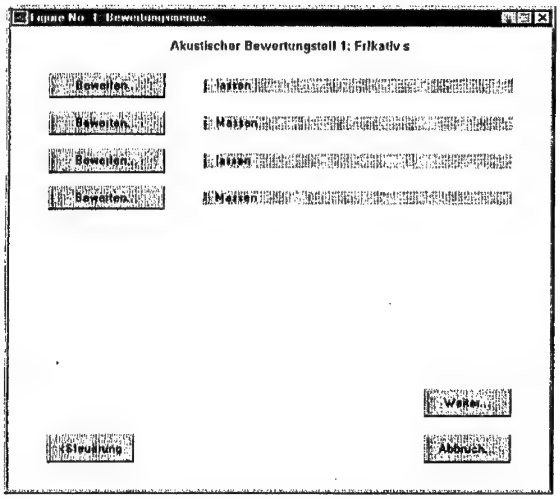

Abbildung 7.5.: Bewertungsoberfläche des Moduls FRIKATIVARTIKULATION. Von hier aus wird die akustische Bewertung jeder Wortaufzeichnung eingeleitet.

entworfene Mustererkenner. Jeder dieser insgesamt 6 Mustererkenner (für weibliche und männliche Sprecher getrennt) wurde in einer zeitaufwendigen Trainingsphase auf die akustischen Eigenschaften der jeweils relevanten Lautpaare anhand einer repräsentativen Stichprobe gesunder Probanden trainiert.
Mustererkenner extrahieren grundsätzlich aus der parametrischen Repräsentation einer Klasse sogenannte Merkmale (Merkmale sind hier geeignete akustische Parameter, die eine möglichst gute Trennbarkeit der beiden Klassen erlauben, also die Information über die relevanten Klassenunterschiede beinhalten) und führen diese einer Klassifikation zu. MODIAS arbeitet hier nach einem Verfahren, wonach im Abstand von 10ms Merkmale extrahiert und anschließend durch eine geeignete Linearkombination auf ein einziges Merkmal, den *Diskriminanz-Score*, abgebildet werden (Merkmalsreduktion). Der in Abbildung 7.6 dargestellte Kurvenverlauf (im folgenden als *Scoreverlauf* bezeichnet) bildet den zeitlichen Verlauf dieses Diskriminanz-Scores ab.
Bedingt durch eine geeignete Wahl der Linearfaktoren, also der spezifischen Gewichtung

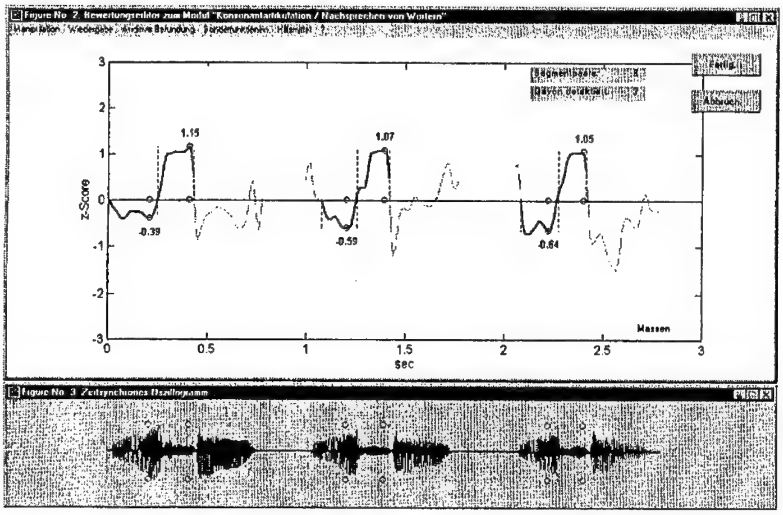

Abbildung 7.6.: Bedienoberfläche des grafischen Bewertungseditors des Moduls FRIKATIVARTIKULATION. Die Kurvenzüge geben Aufschluß über die Erkennungsleistung eines Mustererkenners für die Diskriminanz Frikativ/Vokal.

der Merkmale im Rahmen der Merkmalsreduktion, erreicht der Scoreverlauf bei sprechgesunden Probanden im Mittel

- für den Vokal /a/ jeweils den Wert −1,
- für die Frikative /s/, /ʃ/ bzw. /x/ den Wert +1.

Betragsmäßige Abweichungen nach unten bedeuten eine kontrastärmere Artikulation, Abweichungen nach oben entsprechend eine kontrastreichere Artikulation im Vergleich zur Normgruppe.
Das eigentliche Erkennungsverhalten des Mustererkenners, also z.B. die relative Anzahl der Fehlklassifikationen, rückt hierbei zunehmend in den Hintergrund. Aufgrund der so grundsätzlich verschiedenen akustischen Eigenschaften von Frikativen und Vokalen ist

eine akustisch basierte Klassifikation (durch Mustererkennungsverfahren) auch bei Patienten mit schweren Artikulationsstörungen mit einer so hohen statistischen Sicherheit realisierbar, daß die graduelle Bewertung der Diskriminanz mehr an diagnostisch relevanter Information erwarten läßt als das nicht näher quantifizierbare Urteil der Entscheidungsinstanz.
Durch das Konzept der akustischen Bewertung werden mehrere Ziele verfolgt:

- Grafische Visualisierung der Diskrimination von Frikativ und Vokal durch Projektion eines mit der Güte der artikulatorischen Realisierung korrespondierenden Scoreverlaufs. Hierbei ist die Qualität der Realisierung beider Lautklassen voneinander unabhängig bewert- und interpretierbar.
- Markierung der jeweiligen Extremstellen bezüglich Zeitpunkt und Amplitude innerhalb der Frikativ- bzw. Vokalsegmente .
- Verwertung der markierten Scores von Frikativen und Vokalen für die folgende Auswertung der Fähigkeit zur Kontrastierung Frikativ/Vokal.
- Verwertung der markierten Zeitpunkte der Frikative für die spätere Auswertung der Fähigkeit zur Frikativ-Differenzierung.

Durch den grafischen Bewertungseditor steht dem Untersucher ein Instrument zur Verfügung, mit dessen Hilfe er sowohl den Scoreverlauf betrachten als auch die vom System vorgeschlagenen Markierungen der Extremstellen verändern kann.
Das Menü MANIPULATION gestattet die folgende Auswahl:

- NEUES KONTRAST-/ZIELSEGMENTPAAR SETZEN: Kontrastsegmente korrespondieren gem. der MODIAS-spezifischen Konvention mit Vokalen, Zielsegmente mit Frikativen. Falls nicht vom Algorithmus automatisch erkannt, können diese paarweise gesetzt werden. Hierzu erst den Vokal(kern), dann den Frikativ(kern) mit linker Mousetaste vorgeben. Die zugehörigen Segmentgrenzen werden danach vom Algorithmus automatisch vorgeschlagen.
- SEGMENTKERN VERSCHIEBEN: Diese Funktion gestattet die Verschiebung von Segmentkernen bei unveränderten Segmentgrenzen. Der Ort des Segmentkerns sollte im Idealfall ein relatives oder absolutes Maximum des Scoreverlaufs innerhalb

der zugehörigen Segmentgrenzen markieren. Der korrespondierende Score geht in die weiteren akustischen Analysen ein, daher muß der Segmentkern sorgfältig positioniert sein. Hierzu den zu verschiebenden Segmentkern mit linker Mousetaste selektieren, danach den neuen Ort mit gleicher Taste festlegen.

- SEGMENTPAAR LÖSCHEN: Löschen eines Segmentpaares, bestehend aus dem Segment des Vokals (Kontrastsegment) und dem unmittelbar nachfolgenden Frikativ (Zielsegment) innerhalb eines Wortes. Hierzu Segmentkern des Ziel- oder Kontrastsegmentes eines Wortes selektieren, danach werden automatisch beide entfernt.

Das Menü WIEDERGABE bietet in dem beschriebenen Bewertungseditor umfangreichere Möglichkeiten:

- GESAMT: Die gesamte Aufzeichnung wird mit ihren 3 enthaltenen Einzelwörtern wiedergegeben.
- ÜBERGÄNGE: Die 3 wesentlichen Übergänge, jeweils vom Kontrastvokal auf den folgenden Frikativ, werden innerhalb der detektierten Segmentgrenzen wiedergegeben (/a/ → /s/, /a/ → /ʃ/, /a/ → /x/).
- ZIELSEGMENTE: Wiedergabe der 3 *Zielsegmente* (d.h. der segmentierten Frikative).
- KONTRASTSEGMENTE: Wiedergabe der 3 *Kontrastsegmente* (d.h. der segmentierten Vokale).
- EINZELSEGMENT: Manuelle Wiedergabe einzelner Ziel- bzw. Kontrastsegmente. Hierzu den jeweiligen Segmentkern mit linker Mousetaste selektieren.
- ABSCHNITT VON...BIS: Abschnittsweise Wiedergabe innerhalb definierbarer Grenzen unabhängig von der Segmentierung. Hierzu mit linker Mousetaste Anfang und Ende des wiederzugebenden Abschnittes direkt im Scoreverlauf vorgeben.

Der Menüpunkt HILFSMITTEL enthält eine benutzerdefinierte Zoomfunktion bezüglich der Zeitachse bei fester Skalierung der Ordinate. Die Festlegung des darzustellenden Bereiches erfolgt durch Aufziehen eines Zoomfensters bei gedrückter linker Mousetaste. Die drei weiteren Auswahlen dieses Menüs steuern das Ein- bzw. Ausblenden des

Oszillogrammes sowie die Anpassung der Oszillogramm-Darstellung auf bereits erfolgte Manipulationen oder geänderte Skalierungen.
Der Bewertungseditor wird nach abgeschlossenem Bewertungsvorgang mit dem Button FERTIG verlassen, erst danach werden sämtliche Segmentierungsinformationen gespeichert.
Sind alle Aufzeichnungen der insgesamt 3 akustischen Bewertungsteile bearbeitet, gelangt man durch Betätigung des Buttons WEITER in die Auswertungsoberfläche (Fortsetzung unter 7.2.3 ab Seite 209).

7.2.2.2. Hörersitzung (auditive Analyse)

Wurde innerhalb der Steuerungsoberfläche gem. Abbildung 7.1 der Button HÖRERSITZUNG (auditive Analyse) gewählt, öffnet sich ein Eingabefeld für den Namen des Hörers. An dieser Stelle kann ein beliebiger Text eingegeben werden ohne die stringenten Vorgaben wie bei der Eingabe des Patientenkürzels (z.B. *Dr. Müller im zweiten Versuch*).
Nach Betätigung des Buttons WEITER gelangt man zu der ersten der bis zu 3 aufeinanderfolgenden Befundungsoberflächen (*Auditiver Bewertungsteil*) gem. Abbildung 7.7.
Der erste Schritt ist hier zunächst die Überprüfung der Aussteuerung über den unten links angeordneten Button AUSSTEUERUNGS-CHECK. Nach Betätigung wird dabei ein kurzer Signalton an die angeschlossenen Lautsprecher ausgegeben, der die gleiche Signalaussteuerung wie alle nun folgenden akustischen Einzelwortwiedergaben besitzt.
Das in grüner Schrift eingeblendete Informationsfeld zeigt hier gegebenenfalls an, daß bereits eine Hörersitzung zu den Aufzeichnungen dieser Sitzung stattgefunden hatte. MODIAS gestattet hier bis zu 5 unabhängige Hörersitzungen pro durchgeführter Patientensitzung.

In jeder der Befundungsoberflächen werden maximal 12 Einzelwörter randomisiert dargeboten, wobei der Hörer aus der Darstellung nicht entnehmen kann, um welchen Wort- bzw. Frikativtyp es sich handelt. Nach Betätigung des jeweiligen Buttons OK wird zunächst ein kurzer Signalton als Hinweis auf die folgende Diskriminationsaufgabe ausgegeben, unmittelbar im Anschluß dann das zu prüfende Einzelwort. Der Hörer entscheidet sich nach seinem subjektiven Höreindruck für einen der 3 Frikativtypen /s/, /ʃ/ bzw. /x/und setzt mit der linken Mousetaste ein Kreuz bei dem entsprechenden

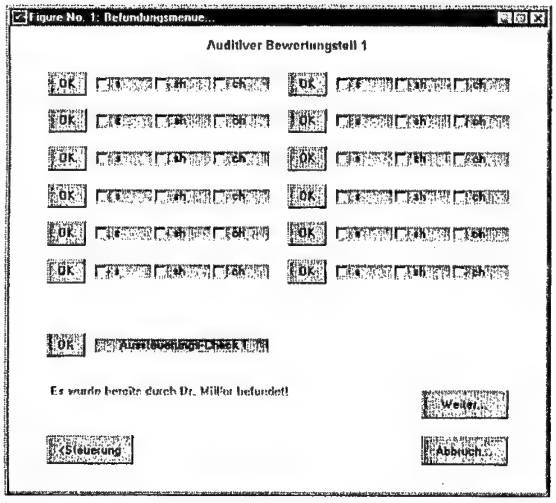

Abbildung 7.7.: Befundungsoberfläche des Moduls FRIKATIVARTIKULATION. Der Hörer iniiert mit der Taste OK die einmalige Wiedergabe eines Einzelwortes und entscheidet sich anschließend für einen (erkannten) Frikativ. Die Reihenfolge der Wiedergaben bezüglich der enthaltenen Frikative ist randomisiert. Bei 36 realisierten Einzelwörtern werden 3 solcher Oberflächen nacheinander bearbeitet.

Feld. Um dabei den unerwünschten Effekt des Einhörens so gering wie möglich zu halten und gleichzeitig das praxisrelevante Spontanurteil des Hörers zu gewinnen, erfolgt die Wiedergabe jeweils nur einmalig. Erst wenn alle Einzelwörter bzw. Diskriminationsaufgaben bearbeitet sind, gelangt man über den Button WEITER schließlich zur Auswertung.

7.2.3. Auswertung, Ergebnisdarstellung und Datenexport

Die Auswertung des Moduls FRIKATIVARTIKULATION widmet sich, wie bereits dargestellt, sowohl der Fähigkeit des Patienten zur Kontrastierung zwischen Frikativen und Vokalen sowie der Fähigkeit zur hinreichenden Frikativ-Differenzierung im Sinne

der Sprachverständlichkeit.

Die Frikativ-Differenzierung kann über zwei Entscheidungsinstanzen (Mensch bzw. Maschine) geprüft werden, wobei die jeweiligen Ergebnisdarstellungen vollständig analog strukturiert und damit unmittelbar vergleichbar gehalten sind.
Die Auswertungsoberfläche gem. Abbildung 7.8, von der aus sämtliche Auswertungen gestartet werden können, teilt sich in die folgenden Bereiche auf:

- Akustische Analysen durch die Entscheidungsinstanz „Maschine"(sowohl Kontrastierung Frikativ/Vokal als auch Frikativ-Differenzierung),
- Auditive Analyse durch die Entscheidungsinstanz „Mensch"(nur Frikativ-Differenzierung).

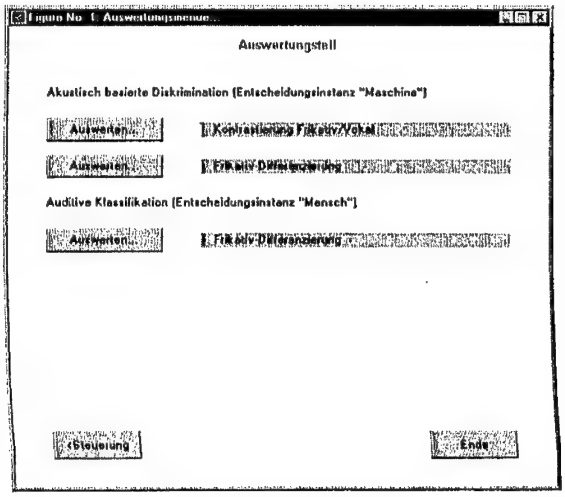

Abbildung 7.8.: Auswertungsoberfläche des Moduls FRIKATIVARTIKULATION. Von hier aus werden die akustischen Analysen (Entscheider: Maschine) oder die auditive Analyse (Entscheider: Mensch) gestartet.

7.2.3.1. Akustische Analyse I: Kontrastierung Frikativ/Vokal

Das Auswertungsblatt gem. Abbildung 7.9 zeigt in der oberen Hälfte getrennt nach Frikativtypen die MITTLEREN SCORES der Frikative (rot bzw. die jeweils rechten Säulen) und der zugehörigen Kontrastvokale (blau bzw. die jeweils linken Säulen). Sie wurden im Rahmen der akustischen Bewertung innerhalb der Bedienungsoberfläche des grafischen Bewertungseditors für jedes einzelne Frikativ- bzw. Vokalsegment ermittelt. Wie bereits dargestellt, sollten diese mittleren Scores bei einem sprechgesunden Probanden betragsmäßig gleich groß sein, aber mit jeweils umgekehrtem Vorzeichen.

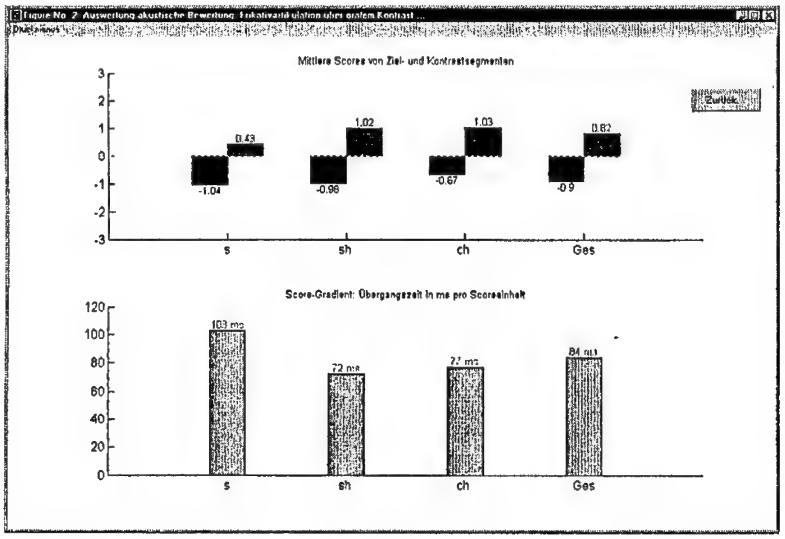

Abbildung 7.9.: Auswertung zum Modul FRIKATIVARTIKULATION, Akustische Analyse zur Kontrastierung Frikativ/Vokal, Blatt 1.

Die Scores sind für die beiden Lautklassen Frikativ bzw. Vokal voneinander unabhängig interpretierbar. Der Betrag -1 entspricht dem Durchschnittswert der zugrunde liegenden Normstichprobe sprechgesunder Menschen für den Vokal /a/, der Betrag +1 dem entsprechenden Durchschnittswert für den jeweiligen Frikativ /s/, /ʃ/ bzw. /x/ (Die Normdatenerhebung wurde anhand des identischen Wortmaterials vollzogen, also in ei-

ner identischen Lautumgebung für die hier betrachteten Lautklassen).
Günstig ist hier eine möglichst hohe Differenz, die ein sensitives Maß für die Fähigkeit des Patienten zur artikulatorischen Kontrastierung darstellt.
In der unteren Hälfte sind die MITTLEREN SCOREGRADIENTEN aufgetragen. Diese Maßzahl ermittelt sich aus den Flankensteilheiten der Scoreverläufe (siehe Kurvenverlauf im Bewertungseditor in Abbildung 7.6 auf Seite 205 bzw. Abbildung 7.10), die sich durch den mehr oder weniger schnellen Übergang vom Vokal zum nachfolgenden Frikativ einstellen. Sie parametriert damit die Fähigkeit des Patienten zu einer schnellen Stellungsänderung der Artikulatoren.

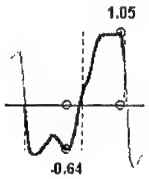

Abbildung 7.10.: Scoreverlauf am Übergang vom Vokal zum Frikativ. Je steiler die Kurve ist, desto schneller wird der artikulatorische Kontrast vom Patienten realisiert.

Das Menü DRUCKEN stellt diverse Varianten der Druckausgabe zur Verfügung wie Farbdruck, Schwarz-Weiß-Druck oder Kopie in die Zwischenablage für die weitere Verwendung in klinischen Berichten usw.

7.2.3.2. Akustische Analyse II: Frikativ-Differenzierung

Dieser Auswertungsteil zeigt das Erkennungsverhalten eines entsprechenden Mustererkenners auf, der allein auf Basis der objektiv meßbaren, akustischen Signaleigenschaften eine Zuordnung zu den 3 Frikativklassen /*s*/, /*ʃ*/ und /*x*/vornimmt. Da er zuvor auf die akustischen Signaleigenschaften einer hinreichend großen Anzahl sprechgesunder Probanden trainiert wurde, spiegeln sich die zu erwartenden Signalveränderungen sprechgestörter Probanden in einer erhöhten Anzahl von Fehlklassifikationen wider.
Im ersten Auswertungsblatt gem. Abbildung 7.11 sind in der oberen Hälfte Säulendia-

gramme für die 3 betrachteten Frikativtypen vorgesehen, die im Verhältnis der Klassifikationsergebnisse *Richtig erkannt* (grün) bzw. *Falsch erkannt* (rot) farblich abgesetzt sind. Die absoluten Anzahlen der jeweiligen Entscheidungen sind in den zugeordneten Säulenabschnitten eingeblendet.

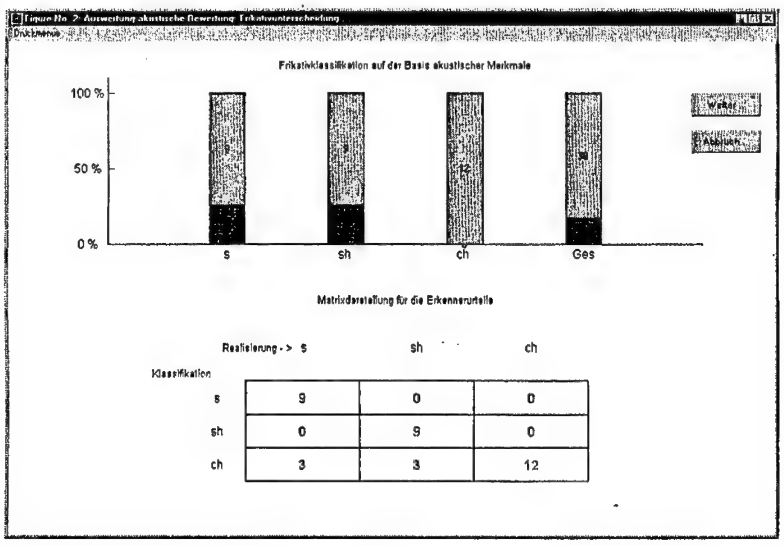

Abbildung 7.11.: Auswertung zum Modul FRIKATIVARTIKULATION, Akustische Analyse der Frikativ-Differenzierung, Blatt 1.

In der unteren Blatthälfte kann das beobachtete Entscheidungsverhalten weiter analysiert werden.

Eine Matrixdarstellung veranschaulicht den Zusammenhang zwischen Zielvorgabe (Was sollte gesprochen werden?) und tatsächlicher Klassifikation (Was wurde vom System erkannt?). Bedingt durch diese Art der Darstellung bilden die korrekten Klassifikationen die Hauptdiagonale, alle Fehlklassifikationen kommen oberhalb bzw. unterhalb dieser Diagonale zu liegen.

Im konkreten Beispiel wurde durch den Mustererkenner der Frikativ /x/ in allen Fällen fehlerfrei erkannt; die Frikative /s/ und /ʃ/ wurden jedoch in jeweils 3 von 12 Fällen fehlklassifiziert als /x/.

Das folgende Auswertungsblatt gem. Abbildung 7.12 bietet dem Untersucher eine Möglichkeit, das dargestellte Entscheidungsverhalten noch eine Ebene tiefer, jetzt unmittelbar in der Entscheidungsebene des Mustererkennersystems zu beurteilen.
Während die für die Analyse der Kontrastierung Frikativ/Vokal (2-Klassen-Diskriminanz) entwickelten Mustererkenner vor dem eigentlichen Klassifikationsprozeß jeweils eine Merkmalsreduktion auf einen *1-dimensionalen Score* vorgesehen hatten (Scoreverlauf wurde im Bewertungseditor dargestellt), sieht der Entwurf des hier beschriebenen Systems zur Analyse der Frikativ-Differenzierung (3-Klassen-Diskriminanz) eine Merkmalsreduktion auf einen *2-dimensionalen Score* bzw. ein Score-Paar vor. Damit findet die Klassifikation in einer Entscheidungsebene statt, d.h., jeder realisierte Frikativ wird im Laufe des Erkennungsprozesses auf einen Punkt in einer 2-dimensionalen Ebene projiziert. Sie ist übersichtlich und einfach grafisch darstellbar.

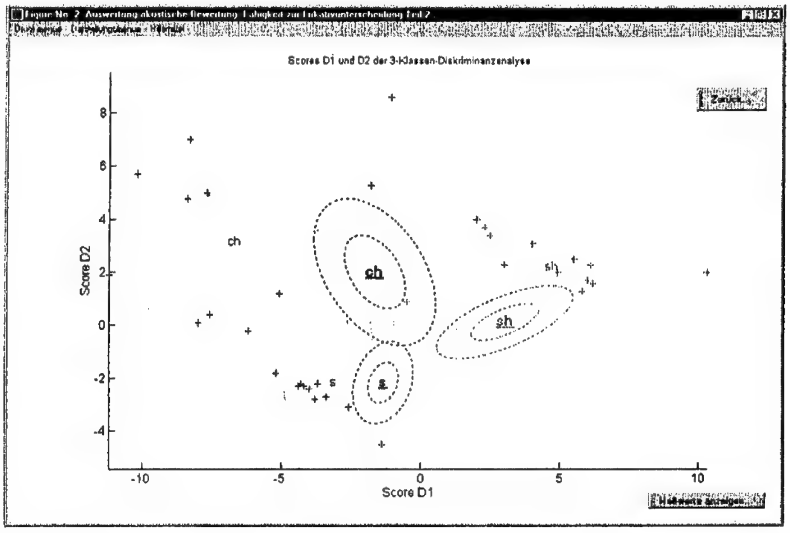

Abbildung 7.12.: Auswertung zum Modul FRIKATIVARTIKULATION, Akustische Analyse der Frikativ-Differenzierung, Blatt 2. Der Untersucher hat hier die Möglichkeit, das Klassifikationsverhalten des realisierten Musterkenners direkt in dessen Entscheidungsebene zu prüfen.

Gestrichelt bzw. unterstrichen eingezeichnet sind die 1-σ und 2-σ-Ellipsen der Normgruppe mit ihren jeweiligen Gruppenzentroiden. Jeder realisierte Frikativ des gerade betrachteten Patienten wird dabei in diese Entscheidungsebene eingeblendet. Durch Vergleich mit den Daten der Normgruppe ist es jetzt für den Untersucher auf einen Blick möglich,

- den Grad der Abweichung von der Norm zu erkennen sowie
- die Auswirkung einer Abweichung tendenziell zu erfassen.

Die Abbildung 7.12 zeigt in Fortsetzung des erwähnten Beispiels sofort die Tendenzen der Frikative /*s*/ und /$\int$/ zum Frikativ /*x*/ an, die zu den festgestellten Fehlklassifikationen führen mußten.
Bedingt durch die zu Grunde liegende Klassifikationsstrategie des Mustererkenners erfolgt die Zuordnung eines Musters nicht umbedingt zu der Klasse mit dem in der Entscheidungsebene nächstgelegenen Gruppenzentroid, d.h., die Entscheidung wird nicht in der Entscheidungsebene durch Abstandsmaße getroffen. Die Klassifikation erfolgt hier vielmehr in statistischem Sinne optimal, d.h., die Zuordnung erfolgt zu der Klasse mit der höchsten Zugehörigkeitswahrscheinlichkeit. Somit werden auch die statistischen Eigenschaften der jeweiligen Klassen berücksichtigt.
Das Menü DRUCKEN stellt diverse Varianten der Druckausgabe zur Verfügung wie Farbdruck, Schwarz-Weiß-Druck oder Kopie in die Zwischenablage für die weitere Verwendung in klinischen Berichten usw.
Das Menü DARSTELLUNG gestattet die Reduzierung der grafischen Darstellung auf die aktuellen Scores jeweils nur eines Frikativtyps mit den zugehörigen σ-Ellipsen bzw. des Zentroids. Damit kann im Einzelfalle die Darstellung durch Ausblendung momentan nicht relevanter Information deutlich an Übersichtlichkeit gewinnen.
Das Menü HILFSMITTEL bietet die Möglichkeit, einzelne Bereiche der Entscheidungsebene durch eine implementierte Funktion ZOOM vergrößert darzustellen. Hierzu kann mit der linken Mousetaste ein entsprechendes Zoomfenster aufgezogen werden; die Rückkehr zur Originaldarstellung erfolgt über die Option REZOOM.

Durch Betätigung des Buttons MESSWERTE ANZEIGEN werden dem Untersucher Zusatzinformationen eingeblendet (Abbildung 7.13), die den Grad und die Richtung von Verschiebungen gegenüber der Normgruppe visualisieren. Im Einzelnen sind dies

- die ermittelten Gruppenzentroide für den aktuell betrachteten Probanden und deren Abstand von den Gruppenzentroiden der Normgruppe in der Entscheidungsebene als Maß für die artikulatorisch bedingten Lautverschiebungen und

- das Verhältnis der durch die Gruppenzentroide aufgepannten Dreiecksflächen von betrachtetem Proband und der Normgruppe als sensitives Maß für die Frikativ-Differenzierung (*Flächenquotient*).

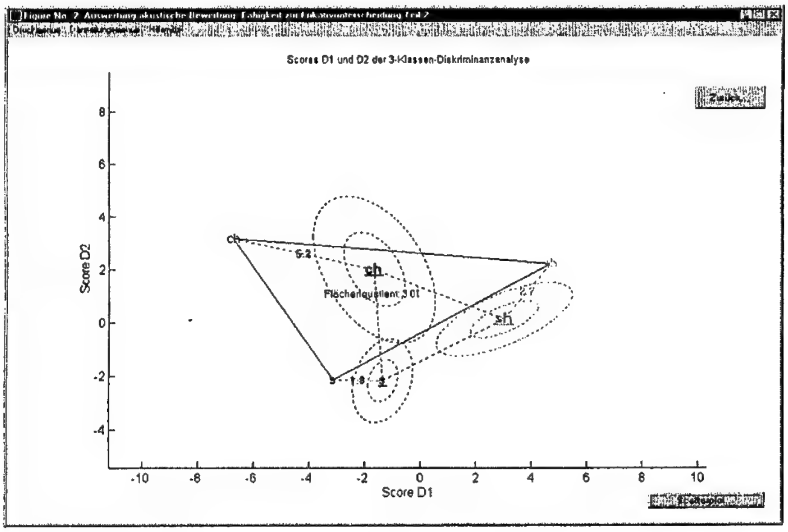

Abbildung 7.13.: Die optionale Einblendung von Zusatzinformationen in die Entscheidungsebene des Mustererkennersystems liefert Aussagen über den Grad der Verschiebung einzelner Frikative wie auch über die Differenzierung der 3 Frikative gegenüber der mittleren Differenzierung der Normgruppe.

Bei Betätigung des Buttons SCATTERPLOT erfolgt die Rückkehr zur ursprünglichen Darstellung ohne Hilfsinformationen gem. Abbildung 7.12.

7.2.3.3. Auditive Analyse: Frikativ-Differenzierung

Während der vorhergehende Abschnitt das Erkennungsverhalten eines trainierten Mustererkenners aufzeigte, wird in diesem Auswertungsteil das Entscheidungsverhalten eines geschulten Hörers dargestellt, der allein auf Basis des subjektiv bewertbaren, perzeptiven Höreindrucks eine Zuordnung zu den 3 Frikativklassen /s/, /ʃ/ und /x/ vornimmt.

Analog zu einem Mustererkenner, der in einem zeitlich befristeten Prozeß auf die akustischen Signaleigenschaften sprechgesunder Probanden trainiert wurde, ist der Mensch durch seine lebenslange Erfahrung auf den Höreindruck gesunder bzw. unauffälliger Sprecher trainiert. Allerdings übertrifft der Mensch mit seiner Erkennungsleistung jede bisher realisierte Maschine um ein Vielfaches, besonders unter dem Aspekt seiner schnellen Adaptionsfähigkeit auf eine weniger kontrastreiche Artikulation (auch mit komplexen Störungsmerkmalen) nach einer nur kurzen Eingewöhnungsphase.

Dieser Adaptionseffekt stellt in der auditiven Befundung von Sprechstörungen durch die resultierende hohe Varianz der Hörerurteile einen erheblichen Nachteil dar. Allerdings muß sich das bereits vorgestellte Instrument der akustischen Analyse erst in der Praxis gegenüber der auditiven Analyse bewähren, so daß auf letztere zunächst noch nicht verzichtet werden darf.

Das erste Auswertungsblatt ist analog zu dem des vorangehenden Abschnittes aufgebaut (siehe Abbildung 7.11), um jetzt auch, neben dem auditiven Befund, einen direkten Vergleich zwischen den beiden Entscheidungsinstanzen *Mensch* und *Maschine* zu ermöglichen. Da das Untersuchungsprotokoll jedoch die unabhängige Befundung mehrerer Hörer zuläßt (bis zu 5), werden hier die kumulierten Urteile sämtlicher Hörer eingeblendet.

Das abschließende Auswertungsblatt gem. Abbildung 7.14 gestattet eine Analyse der hörerspezifischen Urteilsverhalten. Für jeden der bis zu 5 Hörer wird hier pro Frikativtyp jeweils eine farbige Säule zur Visualisierung der betreffenden Fehlklassifikationsraten dargestellt. Unregelmäßigkeiten in der auditiven Befundung können hier optisch sehr leicht erkannt und differenziert betrachtet werden.

Die letzte Zeile der Säulengrafiken stellt ergänzend wieder die Gesamtmittelwerte dar, um z.B. schnell einen Überblick darüber zu erhalten, welcher Hörer in seinem Befund stark von dem Durchschnitt abweicht.

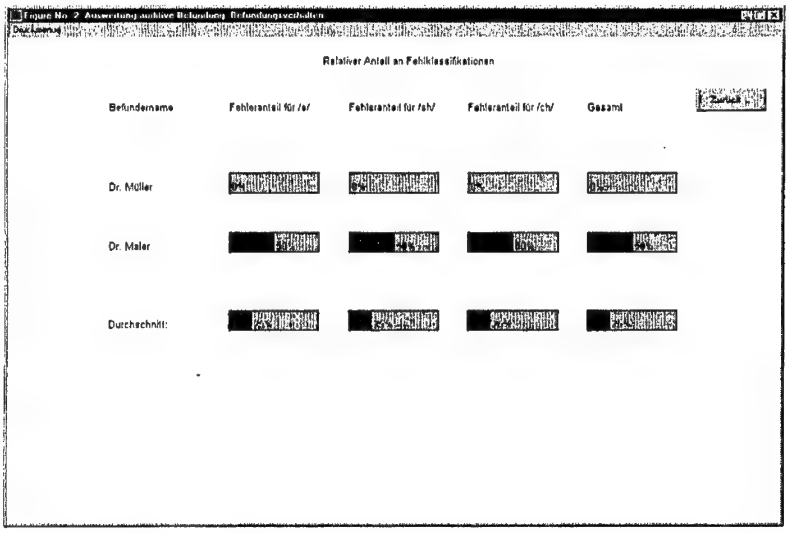

Abbildung 7.14.: Auswertung zum Modul FRIKATIVARTIKULATION, Auditive Analyse der Frikativ-Differenzierung, Blatt 2.

7.3. Technische Realisierung: Entwickelte Verfahren und Algorithmen

7.3.1. Mustererkennerentwurf zur Klassifikation Frikativ/Vokal

7.3.1.1. Methodischer Ansatz

Der erste zu realisierende Mustererkenner soll die beiden Klassen *Frikativ* und *Vokal* anhand ihrer akustischen Eigenschaften differenzieren. Von Interesse ist hier aber weniger die Zuordnung entsprechender Lautsegmente in eine der beiden Klassen, da diese Klassifikationsaufgabe auf Grund der sich stark unterscheidenden akustischen Eigenschaften auch bei gestörter Artikulation noch sehr zuverlässig realisierbar ist. Die Beurteilung allein anhand der meßbaren Fehlkassifikationsrate ließe sich also hierzu nicht sinnvoll heranziehen.
Aus diagnostischer Sicht ist vielmehr ein möglichst sensitiver akustischer Parameter erwünscht, der hoch mit der jeweiligen Artikulationsgüte korreliert und damit sowohl

- die absolute Beurteilung der Artikulation jeder der beiden Lautklassen einzeln, aber auch
- die relative Beurteilung der Artikulation der beiden Lautklassen zueinander zuläßt.

Die letztgenannte Variante führt zur Ableitung eines Differenz- bzw. Kontrastmaßes. Durch die Wahl des Wortmaterials ist der Kontrastlaut, d.h. der Vokal, bei allen Einzelwörtern der beiden Wort-Ensembles gleich; damit kann der Übergang zum nachfolgenden Frikativ als näherungsweise identisch angenommen werden.
Der Entwurf des zu realisierenden Mustererkenners läßt sich gem. Abbildung 7.15 systematisch in Funktionsblöcke untergliedern.

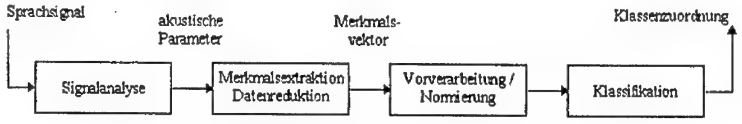

Abbildung 7.15.: Funktionale Blöcke eines Mustererkennersystems.

7. *Realisierung des Moduls FRIKATIVARTIKULATION*

Für den Entwurf eines Mustererkennersystems kommt der Merkmalsextraktion und der Vorverarbeitung eine zentrale Bedeutung zu. Diese Verarbeitungsstufen stellen der nachfolgenden Klassifikation diejenigen Größen zur Verfügung, die für den Entscheidungsprozeß besonders gut geeignet sind. Im Idealfalle sollten hier nur solche Merkmale angeboten werden, die für die Klassentrennung relevant sind.
Für die Lösung des gestellten Klassifikationsproblems Frikativ/Vokal werden in der Signalanalysestufe aus einer geeigneten Spektraldarstellung des Sprachsignals zunächst akustische Parameter abgeleitet und zu *Datenvektoren* zusammengefaßt, die möglichst viele der für die jeweilige Artikulation charakteristischen Eigenschaften beinhalten und damit die klassenspezifischen Unterschiede möglichst gut reproduzieren.
Diese dürfen dann aber nicht von Stimmparametern, wie z.B. Sprechstimmlage, Stimmqualität oder Anregungsintensität, beeinflußt werden. Sie sollen weiterhin auch unabhängig vom gewählten Aufzeichnungssystem und seinen parasitären Einflüssen sein, also z.B. von den Frequenzgängen von Mikrophon bzw. Soundkarte, oder von den stets vorhandenen rechnerinternen Störstrahlungen durch das PC-Netzteil oder die Grafikkarte. Weiterhin sollten die Datenvektoren innerhalb der Zeitfenster, die hinsichtlich artikulatorischer Parameter als stationär anzusehen sind, nur eine geringe Varianz aufweisen.
Die anschließende Merkmalsextraktionsstufe gewinnt aus den bereitgestellten Datenvektoren die Merkmale, die zur Klassentrennung geeignet sind. Die Aufgabe wurde in diesem Teil des MODIAS-Moduls durch Linearkombination der Datenvektoren mit einer speziellen *Diskriminanzfunktion* gleicher Elementzahl gelöst; hierdurch wird jeder mehrdimensionale Datenvektor auf einen diskreten Zahlenwert, den sogenannten Diskriminanzscore D abgebildet.
Diese Scores werden als eindimensionale *Merkmalsvektoren* interpretiert, die

- die gesamte enthaltene Information der Datenvektoren enthalten, dabei aber
- statistische Abhängigkeiten der Vektorelemente entfernen und gleichzeitig eine
- spezifische Gewichtung der Vektorelemente entsprechend ihrer Relevanz für eine Klassentrennung berücksichtigen.

Die Diskriminanzfunktion wird für eine Normstichprobe sprechgesunder Probanden über das statistische Verfahren der *2-Klassen-Diskriminanzanalyse* berechnet und auf eine optimale Klassentrennbarkeit hin optimiert. Dieser Prozeß beinhaltet das stets erforderliche

Training des zu Grunde liegenden Mustererkenners.
Der eigentliche Klassifikationsprozeß wurde mit einem Bayes-Klassifikator realisiert, der über einen funktionalen Zusammenhang für jeden möglichen Punkt im Merkmalsraum angibt, wie wahrscheinlich die Zugehörigkeit des gerade betrachteten Musters zur Klasse der Frikative bzw. zur Klasse der Vokale ist. Damit entsteht über die Zugehörigkeitswahrscheinlichkeit zur tatsächlichen realisierten Klasse ein Parameter, der durch seinen direkten Bezug zur Normgruppe als Maß für die artikulatorische Realisierungsgüte interpretiert werden kann. Die letztlich getroffene Entscheidung des Mustererkenners für die wahrscheinlichste Klasse erfolgt in statistischem Sinne optimal (Optimalklassifikation nach Bayes).

7.3.1.2. Signalanalyse

Es werden ausschließlich akustische Parameter aus den entsprechenden Kurzzeit-Fourierspektren abgeleitet, da hier schon bei oberflächlicher Betrachtung deutliche Unterschiede zwischen Frikativen und Vokalen erkennbar sind (siehe Abbildung 7.16 im Vergleich zu Abbildung 7.17).
Vokalspektren besitzen durch die vollständige Öffnung des Vokaltrakts insgesamt eine hohe Energie, sie zeigen hohe spektrale Anteile im unteren Frequenzbereich von 200 bis ca. 3500 Hz mit den vokalspezifischen Formantfrequenzen.

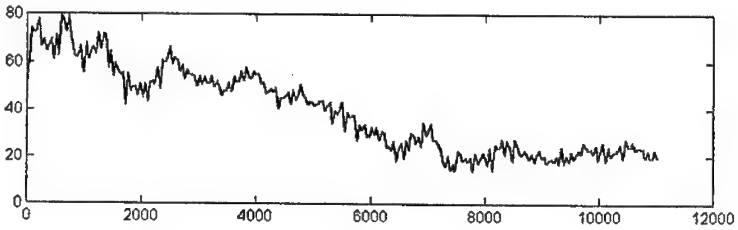

Abbildung 7.16.: Typisches Kurzzeit-Leistungsdichtespektrum des Vokals /a/ mit den charakteristischen Formantfrequenzen (hier bei 700 Hz und 1100 Hz).

Frikative zeigen im Gegensatz hierzu markante spektrale Anteile erst im Frequenzbereich oberhalb von etwa 3 kHz (siehe Abbildung 7.17) bei einer insgesamt aber deutlich niedrigeren Energie gegenüber den Vokalen.

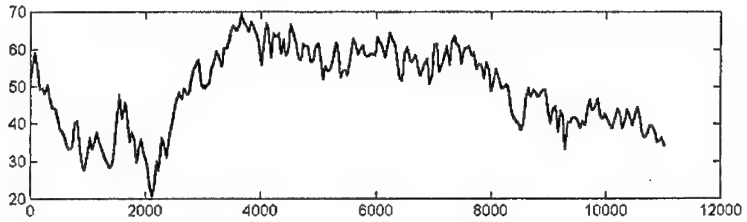

Abbildung 7.17.: Typisches Kurzzeit-Leistungsdichtespektrum des Frikativs /ʃ/.

Schätzung von Kurzzeit-Leistungsdichtespektren: Ein Verfahren zur Schätzung des Leistungsdichtespektrums eines stochastischen Prozesses basiert im Prinzip auf Anwendung der diskreten Fouriertransformation (üblicherweise berechnet nach dem Fast-Fourier-Transformationsalgorithmus FFT) und anschließender Quadrierung der resultierenden spektralen Amplitudendichte. Wie bei jedem Kurzzeitanalyseverfahren muß dabei durch Anwendung einer geeigneten Window-Funktion ein zeitlich begrenztes Signalsegment aus dem Sprachsignal entnommen werden, wodurch die Schätzung des Leistungsdichtespektrums jedoch zunächst verfälscht wird.
Die Anwendung einer Fenster-Funktion im Zeitbereich (d.h. Multiplikation von Sprachsignal und Fenster) korrespondiert im Frequenzbereich mit einer Faltung der jeweiligen Spektraldarstellungen, d.h., das so geschätzte Leistungsdichtespektrum $\hat{P}(\omega)$ besteht aus der Faltung des tatsächlichen Leistungsdichtespektrums $P(\omega)$ mit der quadrierten Fouriertransformierten W der Fenster-Funktion w.

$$\mathcal{E}\{\hat{P}(\omega)\} = \frac{1}{2\pi||w||^2} \int_{-\pi}^{\pi} P(\theta)|W(\omega-\theta)|^2 d\theta \tag{7.1}$$

Der Skalierungsfaktor $||w||^2$ errechnet sich aus der Summe der Quadrate aller Elemente der Fenster-Funktion.

$$||w||^2 = \sum_N w(n)^2 \tag{7.2}$$

Wird dieser Skalierungsfaktor kompensiert, entsteht eine bessere Schätzung des Leistungsdichtespektrums, da sich die Schätzung mit Erhöhung der Segmentlänge an das tatsächliche Leistungsdichtespektrum zunehmend annähert. Bei Anwendung eines nichtrechteckförmigen Fensters (z.B. vom Hamming-Typ), spricht man in diesem Zusammenhang von einem *modifizierten Periodogramm* (z.B. in [Kra94]).

Für MODIAS wird im ersten Schritt das gesamte mit einer Samplingrate von 22050 Hz und einer Auflösung von 16 Bit aufgezeichnete Sprachsignal in überlappende Segmente zu je 512 Samples eingeteilt. Die Überlappung wurde so gewählt, daß alle 10 ms ein jeweils neues Segment beginnt.
Für jedes der entstandenen Segmente wird ein modifiziertes Periodogramm nach der dargestellten Berechnungsmethode unter Nutzung von Hamming-Fenstern ermittelt (vgl. auch Abbildung 4.19 auf Seite 75 und die zugehörige Gleichung 4.9 auf Seite 74).

Rauschnormierung: Die für jedes Segment berechneten modifizierten Periodogramme werden zur späteren Merkmalsextraktion herangezogen. Um dabei Merkmale zu erhalten, die unabhängig von der Kurzzeit-Signalenergie der jeweiligen Segmente sind (im folgenden als *Segmentenergie* bezeichnet), muß vorher eine entsprechende Normierung erfolgen. Ohne diese Normierung würde der Mustererkenner die beiden Klassen Frikativ und Vokal nach entsprechendem Erkennertraining an dem gravierendsten Merkmalsunterschied, also der Segmentenergie, unterscheiden und nicht wie gewünscht an den Unterschieden in den spektralen Verteilungen.
Damit werden aber auch zwangsläufig diejenigen Segmente energienormiert, die gerade in den nicht relevanten Sprechpausen zu liegen kommen. Hierin liegt ein ganz grundsätzliches Problem; denn weicht hier das stets vorhandene Grundrauschen des Aufzeichnungssystems nur geringfügig von einer weißen Rauschcharakteristik ab, was innerhalb einer Störstrahlungsumgebung eines PC stets der Fall sein dürfte, ist das Erkennungsverhalten innerhalb dieser Zeiträume nicht mehr vorhersagbar. Ein niederfrequente Netzstörung würde beispielsweise nach der Normierung des modifizierten Periodogramms sicher zu einer Klassifizierung als Vokal führen, denn die resultierende Form des energienormierten Spektrums ist durch die dominanten Anteile bei relativ niedrigen Frequenzen einem Vokalspektrum wesentlich ähnlicher als einem Frikativspektrum.
Die rauschartigen Störungen werden hier additiv angenommen, daher wird für jedes Segment j bereits zu Anfang der Berechnungen von dem geschätzten Leistungsdichtespektrum (siehe Seite 222) das mittlere Leistungsdichtespektrum der im Untersuchungsprotokoll eigens hierzu vorgesehenen Stilleaufzeichnung subtrahiert und damit dem beschriebenen Effekt vorgebeugt.

7.3.1.3. Merkmalsextraktion und -reduktion

Für die weitere Betrachtung ist es geboten, sich an der Leistungsfähigkeit vergleichbarer Mustererkennungssysteme der Natur zu orientieren. Zur Klärung der Frage, *wieviel* spektrale Information zur Unterscheidung der beiden Laute extrahiert werden muß, sollte sinnvollerweise das menschliche Gehör als biologisches Referenzsystem näher betrachtet werden.

Bildung von gehörrichtigen Frequenzkanälen: Aus psychoakustischen Experimenten weiß man, daß bei der spektralen Auswertung des Gehörs Schallintensitäten innerhalb bestimmter Frequenzbänder aufintegriert werden. Verschiedene spektrale Intensitäten innerhalb dieser Frequenzbänder, die Frequenzgruppen genannt werden, werden vom Gehör zusammengefaßt und als gemeinsame Lautheit wahrgenommen. Das Gehör bildet die Frequenzgruppen an sehr vielen Stellen entlang der Basilarmembran, wobei sich etwas mehr als 600 Frequenzstufen unterscheiden lassen. Die Breite einer Frequenzgruppe auf der Basilarmembran ist dort unabhängig von dem Ort, an dem sie gebildet wird.
Auf einer physikalischen Frequenzskala ergibt sich allerdings eine Abhängigkeit der Breite von der Lage der Frequenzgruppe. Die Breite der Frequenzgruppen beträgt unterhalb 500 Hz etwa 100 Hz, oberhalb von 500 Hz beträgt die Bandbreite etwa 20 Prozent der jeweiligen Mittenfrequenz. Werden die Frequenzgruppen lückenlos nebeneinander auf der Frequenzachse aufgetragen (Abbildung 7.18), so wird der hörbare Frequenzbereich in 24 Frequenzbänder aufgeteilt (vgl. [Zwi61]).
Die Bandgrenzen definieren die psychoakustische Tonheitsskale mit der Dimension Bark. Der funktionale Zusammenhang zwischen Frequenz f und Tonheit z ist z.B. in [Zwi82] wie folgt angegeben:

$$z(f) = 13 \arctan\left(\frac{0.76 \cdot f}{1000}\right) + 3.5 \arctan\left(\frac{f}{7500}\right)^2 \tag{7.3}$$

Der Tonheitsbereich $z = i - 1 \ldots i$ definiert dabei jeweils das Frequenzband Nr. i.
Für alle weiteren Betrachtungen werden nur die Frequenzbänder $1 \ldots 22$ herangezogen. Die Tabelle 7.1 stellt alle 22 (Bark-)Bänder mit ihren gem. Gleichung 7.3 definierten unteren und oberen Frequenzgrenzen orientierend dar.

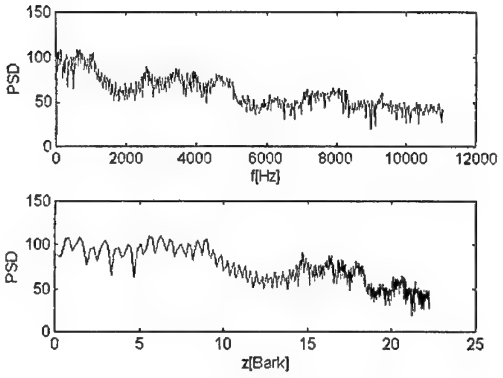

Abbildung 7.18.: Kurzzeit-Leistungsdichtespektrum eines Vokals mit linear skalierter Frequenzachse (oben) und nichtlinear skalierter Tonheitsachse (unten). Die spektrale Auflösung im unteren Frequenzbereich wird dabei deutlich erhöht.

Ableitung der Datenvektoren: Stehen N Abtastwerte eines Signalsegmentes zur Verfügung, wird in der zugehörigen Spektraldarstellung die ursprünglich frequenzkontinuierliche Variable ω durch die frequenzdiskrete Variable ω_m ersetzt. Somit ist das Leistungsdichtespektrum auch nur an den Stützstellen $\omega_m = m\Delta\omega$ bzw. $f_m = m\Delta f$ definiert.
Das Leistungsdichtespektrum besteht bekanntermaßen nach Transformation eines Signalsegments der Länge N aus wiederum N spektralen Komponenten, jedoch sind davon nur $M = N/2$ voneinander unabhängig, so daß sich der relevante Teil des diskreten Leistungsdichtespektrums nur zwischen den Frequenzwerten $0 \cdot \Delta\omega \ldots (N/2-1) \cdot \Delta\omega$ erstreckt.

Das Theorem von PARSEVAL gibt den Zusammenhang zwischen Leistungsdichtespektrum $P(\omega_m)$ und Energie E des betrachteten Signalsegmentes $s(n)$ der Länge N an:

$$E = \frac{1}{M} \sum_{m=0}^{M-1} P(\omega_m) \tag{7.4}$$

Ebenso läßt sich über dieses Theorem auch der für die weitere Betrachtung relevante relative Beitrag eines einzelnen Bark-Bandes i zur (gesamten) Energie E des Segmentes angeben (Er wird in der folgenden Ableitung als E_i bezeichnet, obwohl es sich hier nur noch um einen dimensionslosen Quotienten handelt):

$$E_i = \frac{\sum_{m=u}^{o} P(\omega_m)}{\sum_{m=0}^{M-1} P(\omega_m)} \qquad 0 < u < o < M-1 \tag{7.5}$$

Für MODIAS werden im Rahmen der beschriebenen Realisierung die

- spektralen Komponenten des Leistungsdichtespektrums oberhalb 22 Bark bzw. 10.5 kHz entfernt (vgl. Tabelle 7.1), die
- relativen Energiebeiträge E_i der Bark-Bänder $i = 1 \ldots 22$ gem. Gleichung 7.5 ermittelt und daraus
- pro Segment ein Datenvektor $\mathcal{E}$ gebildet, der sich aus diesen 22 relativen Energiebeiträgen zusammensetzt.

Der Index des betrachteten Segments sei j. Bedingt durch die Art der Normierung gem. Gleichung 7.5 gilt daher:

$$\sum_{i=1}^{22} \mathcal{E}_{ji} = 1 \qquad \forall \qquad j \tag{7.6}$$

Die derart gebildeten Datenvektoren bilden die gehörrichtig skalierte, energienormierte spektrale Energieverteilung der jeweiligen Signalsegmente ab. Bedingt durch die erforderlichen Integrationen über die 22 Frequenzbänder werden hierbei störende Einflüsse durch Stimmparameter wie die absolute Lage der Stimmbandgrundfrequenz und deren Schwankungen oder eventuell auftretende Rauhigkeits- bzw. Behauchtheitsphänomene kompensiert. Sie stellen die Grundlage für alle weiteren Entwurfsschritte des gewählten Mustererkenners dar.
Die folgenden Abbildungen zeigen die typischen, für einen Normsprecher ermittelten Datenvektoren für die 3 Frikative */s/* (Abbildung 7.19), */ʃ/* (Abbildung 7.20) und */x/* (Abbildung 7.21) mit den jeweils zugeordneten Datenvektoren der unmittelbar vorangehenden Kontrastvokale.

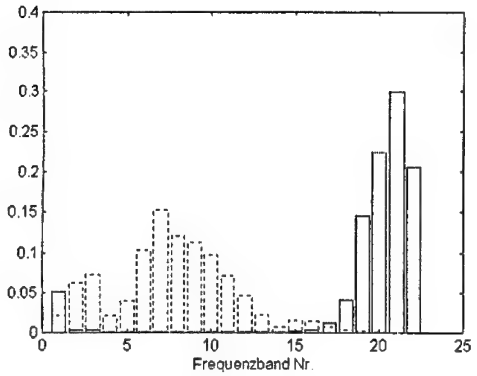

Abbildung 7.19.: Typischer Datenvektor für den Frikativ */s/* (durchgezogene Linie) und den vorangehenden Vokal */a/* (gestrichelte Linie) mit den relativen Energieanteilen der 22 Bark-Bänder

Ableitung der Merkmalsvektoren: Die Abbildungen 7.19, 7.20 und 7.21 zeigen bereits deutlich, daß für einen Mustererkenner die Klassifikation zwischen Frikativ und zugehörigem Kontrastvokal anhand der vorgestellten Datenvektoren vergleichsweise einfach zu realisieren sein dürfte. Schwieriger hingegen muß die im zweiten Teil des Kapitels dargestellte Klassifikation der 3 Frikativklassen selbst eingeschätzt werden, denn die Unterschiede sind hier nur vergleichsweise gering. Jedoch ist beiden Klassifikationsaufgaben gemeinsam, daß die Information über die Klassenzugehörigkeit in der Abhängigkeit der 22 Datenvektorelemente liegt und damit analysiert werden muß.

Die *Diskriminanzanalyse*, erstmals vorgestellt von Sir Ronald Fisher, ist das am häufigsten angewandte Verfahren der Statistik zur Lösung derartiger Problemstellungen [Nor93].
Das zu Grunde liegende Gedankenmodell ist einfach, denn es werden handhabbare Linearkombinationen der unabhängigen Variablen bzw. Prädiktoren gesucht, die eine Basis für eine Zuordnung einzelner Muster in eine der Klassen bilden. Die gesuchte Diskriminanzfunktion besteht aus demjenigen Satz von Koeffizienten, der nach Linearkombination mit allen zur Verfügung gestellten Datenvektoren einer Normgruppe

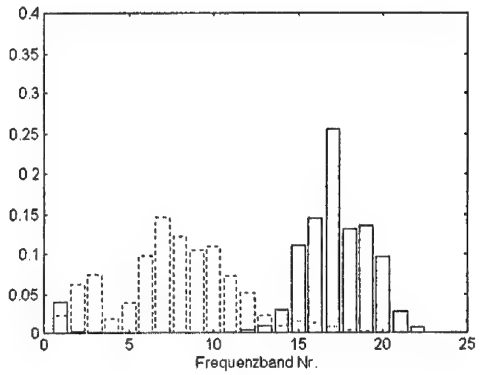

Abbildung 7.20.: Typischer Datenvektor für den Frikativ /ʃ/ (durchgezogene Linie) und den vorangehenden Vokal /a/ (gestrichelte Linie) mit den relativen Energieanteilen der 22 Bark-Bänder

eine optimale Klassentrennbarkeit sicherstellt bzw. eine Zuordnung mit minimaler Fehlklassifikationsrate gestattet. Diese Diskriminanzfunktion besteht hier (wie auch die Datenvektoren) aus 22 Elementen $B_1 \ldots B_{22}$ und einer additiven Konstante B_0.
Das Ergebnis der Linearkombination aus Diskriminanzfunktion und Datenvektor des j-ten Segmentes $\mathcal{E}_j$ ist der Diskriminanzscore D_j:

$$D_j = B_0 + B_1\mathcal{E}_{j_1} + B_2\mathcal{E}_{j_2} + \ldots + B_{22}\mathcal{E}_{j_{22}} = B_0 + \sum_{i=1}^{22} B_i\mathcal{E}_{j_i} \tag{7.7}$$

Damit wird die Gesamtinformation des 22-elementigen Datenvektors $\mathcal{E}_j$ auf einen einzigen Wert D_j reduziert; durch Reduktion entsteht aus dem 22-dimensionalen *Datenvektor* ein 1-dimensionaler *Merkmalsvektor*.

Der Diskriminanzscore wird für jedes Signalsegment $j = 1 : J$ bzw. jeden Datenvektor $\mathcal{E}_j$ der gesamten Aufzeichnung ermittelt. Es entsteht so ein *Scoreverlauf* über der Zeit, der mit der spektralen Verteilung der Segmentspektren und somit mit der Güte der Artikulation korreliert. Üblicherweise weist der Scoreverlauf innerhalb der Realisierungsdauer von Frikativen bzw. Vokalen lokale Extremwerte auf (siehe Abbildung

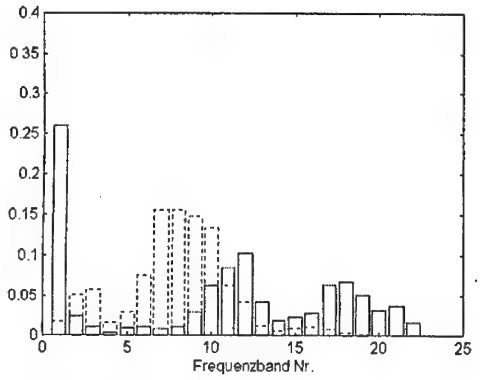

Abbildung 7.21.: Typischer Datenvektor für den Frikativ /x/ (durchgezogene Linie) und den vorangehenden Vokal /a/ (gestrichelte Linie) mit den relativen Energieanteilen der 22 Bark-Bänder

7.22).

Diese Extremwerte des Diskriminanzscores werden für die spätere Klassifikation des Mustererkenners in die beiden Klassen Frikativ bzw. Vokal herangezogen, die hierzu erforderliche Entscheidungsstrategie wird noch eingehend in einem späteren Teil der Arbeit beschrieben. Der aussagekräftigere zeitliche Verlauf des Diskriminanzscores steht dem Untersucher im grafischen Bewertungseditor zur Verfügung (siehe Abbildung 7.6 auf Seite 205).

7.3.1.4. Erkennertraining

Das *Erkennertraining*, also die empirische Optimierung des Mustererkenners zur Lösung der gestellten Klassifikationsaufgabe anhand einer hinreichend großen Trainings-Stichprobe sprechgesunder Probanden, besteht bei der realisierten Entwurfsstrategie in der Durchführung einer 2-Klassen-Diskriminanzanalyse und damit der Gewinnung einer geeigneten Diskriminanzfunktion. Dabei ist für die 3 Klassifikationsaufgaben

- Frikativ /s/ gegenüber Kontrastvokal /a/

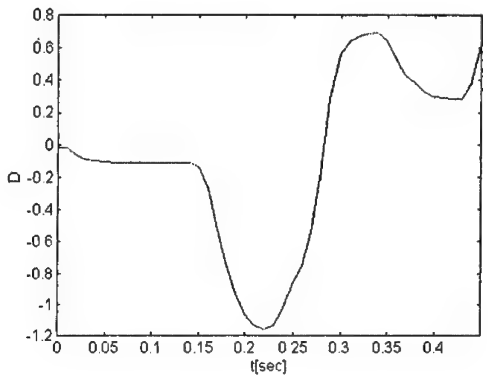

Abbildung 7.22.: Durch Auftragen der Diskriminanzscores D_j über einer Zeitachse entsteht der gesuchte Parameterverlauf, der mit der Güte der Artikulation stark korreliert. Dargestellt ist der Verlauf für das Wort *lassen*, der Diskriminanzscore erreicht hier negative Werte für den Vokal */a/* und positive Werte für den Frikativ */s/*.

- Frikativ /ʃ/ gegenüber Kontrastvokal */a/*
- Frikativ */x/* gegenüber Kontrastvokal */a/*

jeweils eine eigene Diskriminanzfunktion erforderlich. Es entstehen letztlich 3 separate Mustererkenner mit unterschiedlichen Eigenschaften bei identischer Entwurfsstrategie.

Methodischer Ansatz: Die Trainingsphase erfordert eine Handsegmentierung jeder einzelnen Aufzeichnung der Trainings-Stichprobe. Dabei wird über visuelle Prüfung (Oszillogramm) und gleichzeitig perzeptive Kontrolle (Lautsprecher) für jeden Vokal bzw. jeden Frikativ jeweils ein repräsentatives Signalsegment bzw. dessen zugehöriger Datenvektor extrahiert. Die Information bezüglich der Klassenzugehörigkeit ist dabei bekannt. Die Menge der so entnommenen Datenvektoren beider Klassen, jeweils versehen mit der Zusatzinformation der Klasse, bilden die Basis zur Durchführung der Analyse, die im Anschluß mit einem der etablierten Statistik-Programme, hier SPSS, erfolgt.

Am Ende einer 2-Klassen-Diskriminanzanalyse stehen grundsätzlich mehrere Einzelergebnisse zur Verfügung, das sind

- die gesuchte Diskriminanzfunktion, die eine optimale Klassentrennbarkeit anhand des Diskriminanzscores gestattet,
- die beiden Verteilungen der Diskriminanzscores, die sich durch Anwendung der Diskriminanzfunktion auf sämtliche Muster der Trainings-Stichprobe ergeben,
- die Mittelwerte bzw. Zentroide der beiden so entstandenen Verteilungen.

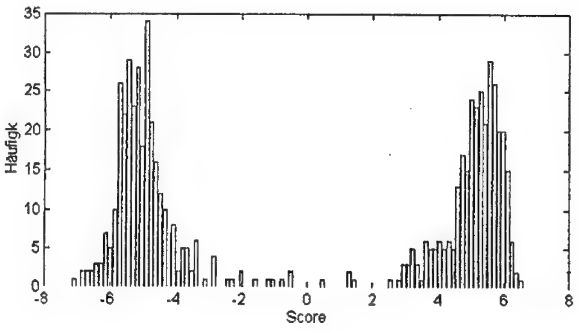

Abbildung 7.23.: Erhaltene Häufigkeitsverteilung der Diskriminanzscores nach erfolgter 2-Klassen-Diskriminanzanalyse. Die Klasse der Vokale besitzt hier negative, die Klasse der Frikative positive Diskriminanzcores.

Das Programmpaket SPSS bestimmt die Diskriminanzfunktion derart, daß die Mittelwerte der beiden entstehenden Verteilungen gleichen Betrag, aber unterschiedliches Vorzeichen besitzen.

Vorläufige Stichprobe: Um für jedes der Einzelwörter jeweils genau einen repräsentativen Datenvektor für den Frikativ und den unmittelbar vorangehenden Kontrastvokal zu erhalten, ist eine manuelle Segmentierung der Aufzeichnungen erforderlich. Wären die Diskriminanzfunktionen auf Basis der Trainings-Stichprobe zu diesem Zeitpunkt bereits

vorhanden, ließe sich die Segmentierung komfortabel in dem grafischen Bewertungseditor (vgl. Abbildung 7.6 auf Seite 205) vornehmen, der den zeitlichen Verlauf der Diskriminanzscores und des zeitsynchronen Oszillogramms darstellt und damit bereits auf die Orte von Frikativen bzw. Vokalen innerhalb des Sprachsignales zeigt. In der Phase des Erkennertrainings standen diese Diskriminanzfunktionen aber logischerweise noch nicht zur Verfügung.
Daher wurden zunächst Diskriminanzfunktionen auf Basis einer vorläufigen Stichprobe berechnet, bestehend aus 1 männlichen und 1 weiblichen Sprecher (Tabelle 7.2) mit jeweils 10 Wiederholungen, um die grafische Segmentierungsunterstützung zumindest eingeschränkt nutzen zu können. Die Ergebnisse dieser Voranalyse sind hier weniger von Interesse und werden daher nicht explizit dargestellt. Zusätzlich wurde innerhalb der Bedienungsoberfläche des grafischen Bewertungseditors eine Spezialroutine implementiert, die

- anhand des vorläufigen Scoreverlaufes die Orte von Frikativ bzw. Vokal vorschlug,
- eine Verschiebbarkeit dieser Orte ermöglichte mit paralleler perzeptiver Kontrolle,
- auf Wunsch die beiden Datenvektoren an den gewählten Zeitpunkten entnahm und automatisch in eine Datenbank portierte.

Trainings-Stichprobe: Die Gewinnung von Trainingsdaten erfolgte an sprechgesunden Probanden im Rahmen einer Erhebung an der Universität der Bundeswehr München und der Entwicklungsgruppe Klinische Neuropsychologie München-Bogenhausen.
Jeder Proband hatte dabei die Aufgabe, sämtliche Einzelwortaufzeichnungen der Patientensitzung des Moduls FRIKATIVARTIKULATION in gewohnter Geschwindigkeit und Stimmlage aufzusprechen. Auf diese Weise entstanden für jeden Probanden Aufzeichnungen von insgesamt 36 Einzelwörtern, davon jeweils 12 für die unterschiedenen Frikative /*s*/, /*ʃ*/ bzw. /*x*/. Vor jeder Aufzeichnung wurden außerdem folgende Zusatzinformationen erfaßt:

- Alter [Jahre]
- Geschlecht [m/w]
- Raucher [Ja/Nein].

Um eventuelle Verfälschungen der Statistik zu vermeiden, wurde vorsorglich jeder Proband befragt, ob nach seinem eigenen, subjektiven Empfinden eine (wenn auch noch so gering ausgeprägte) Stimm- oder Artikulationsstörung vorliegt. Dies wurde von keinem der Probanden bejaht.

Die verwendete Trainings-Stichprobe setzte sich gem. Tabelle 7.3 zusammen. Die im Rahmen der Entwicklung zahlreich durchgeführten Diskriminanzanalysen zeigten zunächst, daß durch Aufteilung der Trainings-Stichprobe nach dem Geschlecht wesentlich bessere Klassifikationsergebnisse erreicht werden konnten. Die Ursache hierfür ist in der signifikant unterschiedlichen Stimmbandgrundfrequenz bzw. Sprechstimmlage begründet. Die männliche Stimme besitzt eine mittlere Grundfrequenz von ca. 120 Hz, die weibliche Stimme von ca. 220 Hz [Cry95]. Dadurch sind systematische Unterschiede in den Teilenergien E_i der Bänder 1-3 (vgl. Gleichung 7.5) und damit in den Datenvektoren zu erwarten.

Das bedeutet, daß geschlechtsspezifische Diskriminanzfunktionen verwendet werden müssen. Die Information bezüglich des Geschlechts wird im Organisationsmenü abgefragt und steht daher im System zur Verfügung.

Auf diese Weise entstanden für jeden der 25 Probanden (Tabelle 7.3) und pro Klassifikationsaufgabe jeweils 12 Datenvektoren für die Klassen Frikativ bzw. Kontrastvokal (insgesamt waren 36 Wörter zu sprechen, davon je 12 pro Frikativtyp bzw. Klassifikationsaufgabe, jedes Wort enthält dabei je einen Frikativ bzw. Kontrastvokal).

Tabelle 7.4 listet sämtliche auf Basis der Trainings-Stichprobe gem. Tabelle 7.3 durchgeführten Diskriminanzanalysen auf mit der jeweils zur Verfügung stehenden Anzahl von Datenvektoren.

Ergebnisse der Diskriminanzanalysen: Die erforderlichen 2-Klassen-Diskriminanzanalysen gem. Tabelle 7.4 wurden mit dem Statistik-Programmpaket SPSS durchgeführt. Dabei resultierten die in Tabelle 7.5 aufgelisteten Diskriminanzfunktionen bzw. deren Koeffizienten $B_0 \ldots B_{22}$. Die zugehörigen Verteilungsmittelwerte (Erläuterung hierzu auf Seite 231) sind in Tabelle 7.6 zusammengestellt.

7.3.1.5. Klassifikation

Aufgabe der Klassifikation ist es nun, nach Anwendung der Diskriminanzfunktion auf ein neues, unbekanntes Muster einer Teststichprobe von dem resultierenden Diskriminanzscore D möglichst sicher auf die Klassenzugehörigkeit zur Klasse k zu schließen. Die Basis für die Konstruktion einer Zuordnungsregel ist die a-priori gegeben Struktur. Unter Ausnutzung der bekannten Klassenzugehörigkeit der Datenvektoren einer Trainings-Stichprobe muß eine Entscheidungsregel definiert werden, die nichts anderes leistet, als die Objekte der Trainings-Stichprobe (und darüber hinaus in Zukunft zuzuordnende Objekte) möglichst fehlerfrei in die vorgegebenen Klassen einzuordnen. Die eigentliche Leistung besteht darin, diese Zuordnung mit den gewählten und zugänglichen Merkmalen zu bewerkstelligen.
Die Konstruktion eines in statistischem Sinne optimalen Klassifikators setzt einen funktionalen Zusammenhang voraus, der für jeden Punkt im Merkmalsraum angibt, wie wahrscheinlich die Zugehörigkeit des Musters zu einer der beiden Klassen ist.
Bei hinreichender Größe der Normstichprobe kann davon ausgegangen werden, daß die Muster einer Klasse von einem Zufallsprozeß erzeugt werden, dessen Gesetzmäßigkeit durch eine gaußförmige Wahrscheinlichkeitsdichtefunktion beschrieben wird.
Die Ableitung einer entsprechenden Entscheidungsregel ist im wesentlichen aus [Rus94] entnommen. Sie erfordert zunächst folgende Definitionen:

- Die Wahrscheinlichkeit, daß ein Muster D zur Klasse k gehört, sei durch die bedingte Wahrscheinlichkeit $p(k|D)$ ausgedrückt.
- Die Kostenfunktion l_{ij} gibt die Kosten an, die entstehen, wenn der Klassifikator sich für die Klasse j entscheidet, obwohl das Muster zur Klasse i gehört.
- Das bedingte Risiko $r_j(D)$ gibt den Erwartungswert der entstehenden Kosten bei einer Entscheidung des Klassifikators zur Klasse j an.

Das bedingte Risiko kann für K Klassen mathematisch wie folgt dargestellt werden:

$$r_j(D) = \sum_{i=1}^{K} l_{ij} \cdot p(i|D) \tag{7.8}$$

Die Entscheidung ist dann optimal, wenn das bedinge Risiko und damit die Kosten minimal sind, also gilt:

$$r_j(D) \rightarrow Min \tag{7.9}$$

Der Ausdruck kann vereinfacht werden, wenn die richtige Entscheidung des Klassifikators keine Kosten, die falsche Entscheidung hingegen konstante Kosten verursacht, also:

$$l_{ij} = 0 \quad \text{bei richtiger Entscheidung und} \quad l_{ij} = C \quad \text{bei falscher Entscheidung} \tag{7.10}$$

Damit ergibt sich für das bedingte Risiko folgender Ausdruck:

$$r_j(D) = \sum_{k=1}^{j-1} C \cdot p(k|D) + \sum_{k=j+1}^{K} C \cdot p(k|D) = \sum_{k=1}^{K} C \cdot p(k|D) - C \cdot p(j|D) \tag{7.11}$$

bzw. nach Vereinfachung

$$r_j(D) = C \cdot \left[\sum_{k=1}^{K} p(k|D) - p(j|D) \right] \tag{7.12}$$

Da das Muster D sicher zu einer der K Klassen gehört, gilt:

$$\sum_{k=1}^{K} p(k|D) = 1 \tag{7.13}$$

Damit vereinfacht sich Gleichung 7.12 zu:

$$r_j(D) = C \cdot [1 - p(j|D)] \tag{7.14}$$

Das Minimum des bedingten Risikos wird gefunden, wenn jeweils diejenige Klasse ausgewählt wird, für die die bedingte Wahrscheinlichkeit $p(k|D)$ maximal ist. Die kostenoptimale Entscheidungsregel lautet damit:

$$D \in i \quad \text{wenn} \quad p(i|D) > p(j|D) \quad \text{für alle } j = 1 \ldots K, j \neq i \tag{7.15}$$

Diese Entscheidungsregel ist als *Identifikationsregel* bekannt, die dem *Bayes-Klassifikator* zu Grunde liegt. Die Gebietseinteilung des vom hier eindimensionalen Diskriminanzscore D aufgespannten Merkmalsraumes läßt sich grafisch anschaulich darstellen (Abbildung 7.24).

Die bedingte Wahrscheinlichkeit $p(k|D)$ gibt Aufschluß darüber, wie wahrscheinlich die

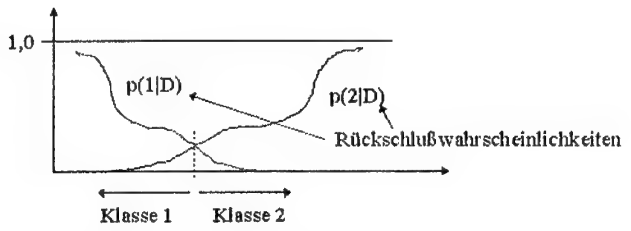

Abbildung 7.24.: Gebietseinteilung des Merkmalsraumes für den Bayes-Klassifikator bei 2 Klassen

Zugehörigkeit eines Musters zur Klasse k bei Beobachtung des Diskriminanzscores D ist (Rückschlußwahrscheinlichkeit). Somit ordnet die Identifikationsregel das beobachtete Muster der statistisch wahrscheinlichsten Klasse zu. Sie kann über den Satz von BAYES berechnet werden, der den folgenden Zusammenhang definiert:

$$P(D|k) = \frac{p(k|D) \cdot p(D)}{p(k)} \tag{7.16}$$

Die jetzt enthaltene bedingte Wahrscheinlichkeit $p(D|k)$ stellt die klassenweise Wahrscheinlichkeitsfunktion bzw. *Likelihood-Funktion* dar, deren Parameter im Gegensatz zur Rückschlußwahrscheinlichkeit durch Beobachtung des stochastischen Prozesses unmittelbar geschätzt werden kann. Sie beschreibt die Wahrscheinlichkeit, daß der beobachtete Diskriminanzscore D von einem Muster der Klasse k erzeugt wird. Die Wahrscheinlichkeit $p(k)$ ist die Auftrittswahrscheinlichkeit der Klasse k, die den Prozeß charakterisiert und a priori bekannt ist.
Durch Anwendung des Satzes von BAYES kann daher die Entscheidungsregel des Bayes-Klassifikators auch umgeformt werden. Sie lautet in anderer Form:

$$D \in i \quad \text{wenn} \quad p(i) \cdot p(D|i) > p(j) \cdot p(D|j) \quad \text{für alle } j = 1 \ldots K, j \neq i \tag{7.17}$$

Hierbei wird eine wichtige Eigenschaft erkennbar, nämlich daß die Entscheidung des Bayes-Klassifikators wesentlich von der Auftrittswahrscheinlichkeit der jeweiligen Klasse abhängt. Die Muster der häufiger auftretenden Klasse werden sicherer klassifiziert als die Muster der seltener auftretenden Klasse. Vom systemtheoretischen Aspekt betrachtet besitzen aber gerade die seltener auftretenden Muster den höheren Informationsgehalt und bedürfen daher einer sichereren Klassifikation.

Um diesen Nachteil des Bayes-Klassifikators zu kompensieren, wird die zur Ableitung der Entscheidungsregel eingeführte Kostenfunktion modifiziert. Die Kosten für eine richtige Entscheidung sollen gleich Null sein, die Kosten für eine Fehlentscheidung sollen sich jetzt umgekehrt proportional zur Auftrittswahrscheinlichkeit der Klasse verhalten.

$l_{ij} = 0$ bei richtiger Entscheidung und $l_{ij} = 1/p(i)$ bei falscher Entscheidung (7.18)

Damit entsteht eine neue Entscheidungsregel, die im Gegensatz zu der vorherigen unabhängig von den klassenspezifischen Auftrittswahrscheinlichkeiten ist.

$$D \in i \quad \text{wenn} \quad p(D|i) > p(D|j) \quad \text{für alle } j = 1 \ldots K, j \neq i \tag{7.19}$$

Die Zuordnung wird jetzt zu derjenigen Klasse getroffen, deren klassenspezifische Wahrscheinlichkeitsfunktion für den beobachteten Diskriminanzscore D den höchsten Wert liefert. Diese Entscheidungsregel definiert den *Maximum-Likelihood-Klassifikator*, dessen Gebietseinteilung des Merkmalsraumes sich auch wieder grafisch veranschaulichen läßt (Abbildung 7.25).

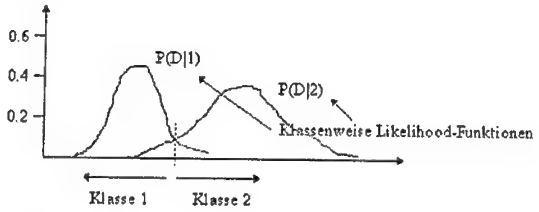

Abbildung 7.25.: Gebietseinteilung des Merkmalsraumes für den Maximum-Likelihood-Klassifikator bei 2 Klassen

Durch Vergleich der beiden Entscheidungsregeln von Bayes- und Maximum-Likelihood-Klassifikator wird klar, daß bei gleichen Auftrittswahrscheinlichkeiten $p(k)$ der betrachteten Klassen k beide Klassifikatoren identisch sind bzw. deren Entscheidungsregeln identische Ergebnisse liefern. Damit ist der Maximum-Likelihood-Klassifikator ein Spezialfall des Bayes-Klassifikators.

Diese Eigenschaft wurde bei der Realisierung in MODIAS genutzt, indem im Rahmen

des Erkennertrainings bzw. der 2-Klassen-Diskriminanzanalyse jeweils genau gleich viele Datenvektoren für die beiden Klassen Frikativ bzw. Vokal herangezogen wurden und damit die gleiche Auftrittswahrscheinlichkeit der beiden Klassen erzwungen wurde. Dieser Ansatz wurde zuvor bereits bei der Auswahl des Wortmaterials berücksichtigt, da jedes Einzelwort je 1 Frikativ bzw. Kontrastvokal enthalten muß. Im Folgenden wird daher die gewählte Realisierung eines Bayes-Klassifikators aufgezeigt, der auf Grund des Spezialfalles gleicher Auftrittswahrscheinlichkeiten zugleich auch einen Maximum-Likelihood-Klassifikator darstellt.

Die Anwendung der Entscheidungsregel gem. Gleichung 7.19 und damit der statistischen Optimalklassifikation bedarf zur Implementierung zunächst der Ermittlung der klassenweisen Wahrscheinlichkeitsfunktionen. Es ist im Allgemeinen nicht möglich, diese Funktionen tatsächlich zu bestimmen.
Als Ausweg bietet sich an, einen bestimmten Typ der Verteilung anzunehmen. Eine solche Annahme kann z.B. aus der Kenntnis über den Prozeß, der die Muster erzeugt, begründet sein. Die Aufgabe beschränkt sich dann auf die Bestimmung des Satzes von Parametern, der die entsprechende Verteilung beschreibt.

Sämtliche Elemente der ursprünglichen 22-elementigen Datenvektoren sind stetige Zufallsvariablen, die einer statistischen Streuung unterliegen. Nach dem zentralen Grenzwerttheorem (z.B. [Bor99]) ist damit insbesondere auch der Diskriminanzscore, der aus einer gewichteten Summation aller dieser Elemente besteht, eine stetig normalverteilte Zufallsgröße (ein hinreichend großer Stichprobenumfang vorausgesetzt!). Daher wurden im Rahmen der Realisierung die gesuchten klassenweisen *Wahrscheinlichkeitsfunktionen* durch gaußförmige *Wahrscheinlichkeitsdichtefunktionen* ersetzt.
Das verwendete Statistik-Programmpaket SPSS setzt implizit solche Normalverteilungen voraus. Der Berechnungsalgorithmus ist zusätzlich noch derart gestaltet, daß die beiden Wahrscheinlichkeitsdichten nach $N(-D,1)$ bzw. $N(D,1)$ normalverteilt sind, d.h. betragsmäßig gleiche Mittelwerte bei einer Standardabweichung von 1 besitzen. Die Annahme von Normalverteilungen läßt sich in indirekter Weise rechtfertigen, wenn damit letztlich eine hohe Erkennungsrate erzielt werden kann ([Rus94]). Die weitere Klassifikationsstrategie fußt auf dieser Annahme.

Die Zuordnung zu einer der beiden Klassen Frikativ bzw. Vokal ist unter den ge-

troffenen Annahmen

- normalverteilter Wahrscheinlichkeitsdichten bei
- betragsmäßig gleichen Verteilungsmittelwerten

einfach. Es ist leicht vorstellbar, daß sich die beiden klassenspezifischen Wahrscheinlichkeitsdichtefunktionen auf Grund ihrer betragsmäßig gleichen Verteilungsmittelwerte stets bei $D = 0$ schneiden. Die hier gewählte Klassifikationsstrategie schreibt ja gerade die Zuordnung zu der Klasse mit dem maximalen Wert der Wahrscheinlichkeitsdichtefunktion vor, damit fällt die Entscheidung des realisierten Mustererkenners ausschließlich anhand des Vorzeichens des Diskriminanzscores. Hierzu muß aus der ursprünglichen Diskriminanzanalyse der Trainings-Stichprobe neben der Diskriminanzfunktion selbst nur noch einer der beiden Verteilungsmittelwerte bekannt sein.

An dieser Stelle der Berechnungen stehen also für jedes beliebige Testmuster zwei Informationen zur Verfügung, das sind

- der erreichte Diskriminanzscore D, sowie
- die zugeordnete Klasse k.

Zusätzlich wurde in die Auswertung des MODIAS-Moduls FRIKATIVARTIKULATION neben der reinen Aussage der Zuordenbarkeit zur Klasse Frikativ bzw. Vokal auch die Wahrscheinlichkeit der Zugehörigkeit zu dieser Klasse berücksichtigt, um die Entscheidung des Mustererkenners transparenter zu machen und damit eine Aussage zu erhalten, mit welcher statistischen Sicherheit diese Klassifikation vorgenommen wurde. Zielrichtung ist auch hier, ein Maß für die artikulatorische Realisierungsqualität zu erhalten, die aber im Gegensatz zum reinen Verlauf des Diskriminanzscores in einem Wertebereich zwischen 0 und 1 besonders anschaulich skaliert ist. Hierfür muß zumindest eine der beiden Rückschlußwahrscheinlichkeiten ermittelt werden.
Für stetige Zufallsgrößen kann die Wahrscheinlichkeit für das Auftreten eines bestimmten Diskriminanzscores nicht mehr angegeben werden. Vielmehr ist nur noch die Wahrscheinlichkeit für das Auftreten von Ereignissen sinnvoll, die sich in einem bestimmten Intervall der Zufallsvariablen befinden.

Daher kann nicht mehr die zur Berechnung der Rückschlußwahrscheinlichkeit erforderliche bedingte Wahrscheinlichkeit für die Beobachtung des spezifischen Diskriminanzscores p(D/k), sondern nur die für p(D<D*/k) bzw. p(D>D*/k) berechnet werden. Damit wird also mit dieser bedingten Wahrscheinlichkeit die Fragestellung beantwortet, wie wahrscheinlich der Diskriminanzscore D* oder ein noch weiter vom Klassenmittelwert entfernt liegender Score auftritt.
An dieser Stelle erfolgt durch Integration der Übergang von den Wahrscheinlichkeitsdichtefunktionen auf die zugehörigen Wahrscheinlichkeitsverteilungen, also der Gauß'schen Fehlerfunktion. Es ergeben sich nach Aufteilung des Merkmalsraumes in 4 Abschnitte die nachfolgenden Ausdrücke für die gesuchten bedingten Wahrscheinlichkeiten bzw. Likelihood-Funktionen (siehe Abbildung 7.26). Die Funktion Φ ist das Gauß'sche Fehlerintegral.
Die Klasse 1 (Vokale) besitze den Verteilungsmittelwert $-D_m$, die Klasse 2 (Frikative) den Verteilungsmittelwert $+D_m$.

$$x < -D_m \text{ (Klasse 1):} \quad p(x < D|1) = \frac{p(x < D, 1)}{p(1)} = 2 \cdot \Phi_1(D) \tag{7.20}$$

$$x > -D_m \text{ (Klasse 1):} \quad p(x > D|1) = \frac{p(x > D, 1)}{p(1)} = 2 \cdot \{1 - \Phi_1(D)\} \tag{7.21}$$

$$x < +D_m \text{ (Klasse 2):} \quad p(x < D|2) = \frac{p(x < D, 2)}{p(2)} = 2 \cdot \Phi_2(D) \tag{7.22}$$

$$x > +D_m \text{ (Klasse 2):} \quad p(x > D|2) = \frac{p(x > D, 2)}{p(2)} = 2 \cdot \{1 - \Phi_2(D)\} \tag{7.23}$$

Bedingt durch die gleichen Auftrittswahrscheinlichkeiten der beiden Klassen und die betragsmäßig gleichen Verteilungsmittelwerte werden beide Kurven auch gleich berechnet, sie sind lediglich gegeneinander verschoben.
Die Umrechnung auf die gesuchten Rückschlußwahrscheinlichkeiten erfolgt wieder über das bekannte Bayes-Theorem. Es reicht hierbei aus, nur eine dieser Wahrscheinlichkeiten (hier: Frikative bzw. Klasse 2) zu berechnen, da sich die Rückschlußwahrscheinlichkeiten beider Klassen stets zu 1 addieren müssen und damit einfach umrechenbar sind.

$$p(2|x < D) = \frac{p(x < D|2) \cdot p(2)}{p(x < D)} = \frac{p(x < D|2) \cdot p(2)}{p(x < D|2) \cdot p(2) + p(x < D|1) \cdot p(1)} \tag{7.24}$$

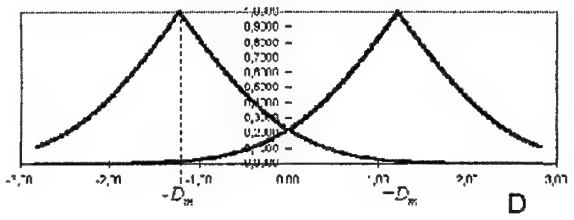

Abbildung 7.26.: Einführung abschnittsweise definierter Likelihood-Funktionen mit Gebietsaufteilung des Merkmalsraumes in 4 getrennte Bereiche.

Damit reduziert sich die Aufteilung des Merkmalsraumes auf 3 Abschnitte:

$$x < -D_m: \quad p(2|x < D) = \frac{\Phi_2(D)}{\Phi_2(D) + \Phi_1(D)} \tag{7.25}$$

$$x > +D_m: \quad p(2|x < D) = \frac{1 - \Phi_2(D)}{1 - \Phi_2(D) + 1 - \Phi_1(D)} \tag{7.26}$$

$$x > -D_m \text{ und } x < +D_m: \quad p(2|x < D) = \frac{\Phi_2(D)}{\Phi_2(D) + 1 - \Phi_1(D)} \tag{7.27}$$

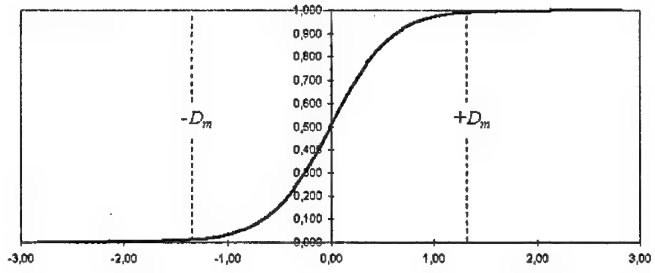

Abbildung 7.27.: Abschnittsweise definierte Rückschlußwahrscheinlichkeit zur Klasse der Frikative in Abhängigkeit des Diskriminanzscores D

7. Realisierung des Moduls FRIKATIVARTIKULATION

Diese abschnittsweise definierte Funktion beschreibt die Wahrscheinlichkeit der Zugehörigkeit zur Klasse der Frikative bei Beobachtung des Diskriminanzscores D. Sie zeigt erwartungsgemäß Punktsymmetrie, steigt streng monoton an und schneidet die Ordinate im Punkt (0,0.5), also der bereits abgeleiteten Entscheidungsschwelle in Abhängigkeit des Vorzeichens des Diskriminanzscores (Abbildung 7.27).
Ist diese Funktion bekannt, kann für jeden erhaltenen Diskriminanzscore eines Testmusters die Zugehörigkeitswahrscheinlichkeit abgelesen werden.

MODIAS liefert diese Wahrscheinlichkeit für die lokalen Extremwerte des Scoreverlaufes, also an den Zentren von Vokalen bzw. Frikativen.

7.3.2. Mustererkennerentwurf zur Klassifikation in 3 Frikativklassen

7.3.2.1. Entwurfsstrategie

Der zweite zu realisierende Mustererkenner soll die 3 Klassen /s/, /ʃ/ und /x/ anhand ihrer akustischen Eigenschaften differenzieren. Während der erste Mustererkenner den Problembereich einer kontrastreichen Artikulation von Frikativen gegenüber Vokalen beleuchtete, rückt jetzt der ebenso relevante Problembereich einer möglichst guten artikulatorischen Unterscheidung der 3 Frikativtypen selbst in den Mittelpunkt des Interesses. Dabei soll die Bewertung ausschließlich anhand des Klassifikationsverhaltens bzw. der meßbaren Fehlklassifikationsrate eines Mustererkenners erfolgen.
Der Entwurf gliedert sich auch hier wieder systematisch in die bereits in Abbildung 7.15 gezeigten funktionalen Blöcke.
Für die Lösung des gestellten Klassifikationsproblems werden in der Signalanalysestufe aus einer geeigneten Spektraldarstellung des Sprachsignals zunächst akustische Parameter abgeleitet und zu Datenvektoren zusammengefaßt, die möglichst viele der für die jeweilige Artikulation charakteristischen Eigenschaften beinhalten und damit die klassenspezifischen Unterschiede hinreichend reproduzieren. Diese dürfen aber nicht von Stimmparametern wie z.B. Sprechstimmlage, Stimmqualität oder Anregungsintensität beeinflußt werden. Sie sollten weiterhin auch unabhängig vom gewählten Aufzeichnungssystem und seinen parasitären Einflüssen sein, also z.B. von den Frequenzgängen von Mikrophon und Soundkarte oder von den stets vorhandenen rechnerinternen Störstrahlungen durch das PC-Netzteil oder die Grafikkarte. Weiterhin sollten die Datenvektoren

innerhalb der Zeitfenster, in denen perzeptiv eine konstante Artikulation vorliegt, nur eine geringe Varianz aufweisen.
Die anschließende Merkmalsextraktionsstufe gewinnt aus den bereitgestellten Datenvektoren die Merkmale, die zur Klassentrennung geeignet sind. Diese Aufgabe wurde in diesem Teil des MODIAS-Moduls durch Linearkombination der Datenvektoren mit *zwei* speziellen Diskriminanzfunktionen gleicher Elementzahl gelöst; hierdurch wird jeder mehrdimensionale Datenvektor auf ein Zahlenpaar, dem Diskriminanzscorepaar $(D1, D2)$ abgebildet. Diese Wertepaare werden als *zweidimensionale Merkmalsvektoren* interpretiert, die

- die gesamte enthaltene Information der Datenvektoren enthalten, dabei aber
- statistische Abhängigkeiten der Vektorelemente entfernen und gleichzeitig eine
- spezifische Gewichtung der Vektorelemente entsprechend ihrer Relevanz für eine Klassentrennung berücksichtigen.

Die beiden Diskriminanzfunktionen werden anhand einer Trainings-Stichprobe sprechgesunder Probanden über das statistische Verfahren der *3-Klassen-Diskriminanzanalyse* berechnet und auf eine optimale Klassentrennbarkeit optimiert. Dieser Prozeß beinhaltet das stets erforderliche Training des zu Grunde liegenden Mustererkenners.
Der eigentliche Klassifikationsprozeß wurde im Rahmen der MODIAS-Implementierung mit einem *Mahalanobis*-Abstandsklassifikator realisiert.

7.3.2.2. Signalanalyse

Es werden ausschließlich akustische Parameter aus den entsprechenden Kurzzeit-Leistungsdichtespektren abgeleitet, da im Spektralbereich Unterschiede zwischen den 3 Frikativen erkennbar sind. Frikativspektren zeigen grundsätzlich wegen der stimmlosen bzw. rauschartigen Anregung und des eingeengten Artikulationstraktes eine vergleichsweise geringe Energie, sie zeigen markante spektrale Anteile erst in einem Frequenzbereich oberhalb von ca. 4 kHz.
Der Frikativ */s/* ist in diesem Frequenzbereich durch einen relativ konstanten Verlauf des Leistungsdichtespektrums charakterisiert (Abbildung 7.28).

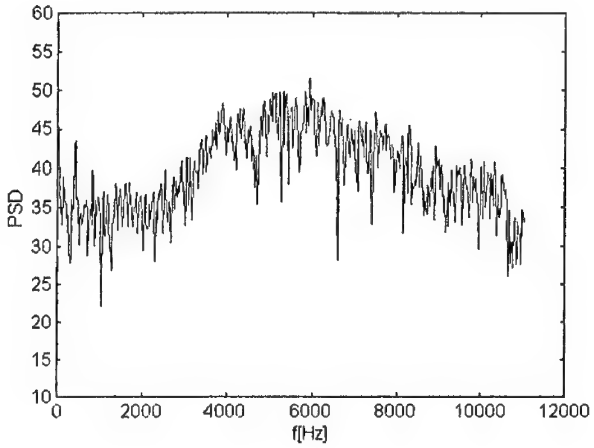

Abbildung 7.28.: Typisches Leistungsdichtespektrum des Frikatives /s/ mit rauschartiger Charakeristik oberhalb einer Frequenz von ca. 4 kHz.

Der Frikativ /ʃ/ zeigt einen zu hohen Frequenzen hin abfallenden Verlauf des Leistungsdichtespektrums, da die bei der Artikulation gerundeten und nach vorne gestülpten Lippen eine tiefpaßartige Filterung bewirken (Abbildung 7.29).
Der Frikativ /x/ hingegen zeigt eine Besonderheit. Während bei dem Frikativ /s/ die Anregung durch das Passieren des aspirierten Luftstromes an der sehr weit vorne liegenden Engstelle zwischen Zunge und vorderem Gaumen entsteht, erfolgt hier die Anregung weiter zurückgesetzt zwischen Zungenrücken und hinterem Teil des Gaumens. Dadurch erfährt das Anregungssignal noch eine Filterung durch den vorderen Teil des Vokaltraktes, das ursprünglich rauschartige Anregungssignal wird gefärbt, und es prägen sich bei niedrigeren Frequenzen die eigentlich vokaltypischen Formanten aus (Abbildung 7.30).

Schätzung von Kurzzeit-Leistungsdichtespektren: Die Schätzung von Kurzzeit-Leistungsdichtespektren erfolgt identisch zu Kapitel 7.3.1.2. Daher sei an dieser Stelle auf Seite 222 verwiesen.

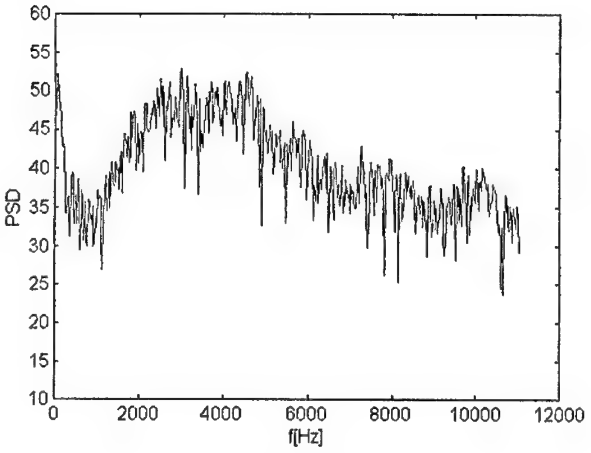

Abbildung 7.29.: Typisches Leistungsdichtespektrum des Frikatives /ʃ/. Durch die Vorstülpung der Lippen entsteht eine tiefpaßartige Bedämpfung höherer Frequenzen.

Rauschnormierung: Die Rauschnormierung erfolgt identisch zu Kapitel 7.3.1.2. Daher sei an dieser Stelle auf Seite 223 verwiesen.

7.3.2.3. Merkmalsextraktion und -reduktion

Bildung von gehörrichtigen Frequenzkanälen: Die Bildung gehörrichtiger Frequenzkanäle erfolgt identisch zu Kapitel 7.3.1.3. Daher sei an dieser Stelle auf die Seite 224 verwiesen.

Ableitung der Datenvektoren: Die Frikativ-Datenvektoren gem. Abbildung 7.19, 7.20 und 7.21 zeigen, daß für einen Mustererkenner die Klassifikation zwischen den 3 Frikativklassen möglich sein muß, wenn auch schwieriger als bei der Unterscheidung Frikativ gegenüber Vokal. Auch bei der vorliegenden Klassifikationsaufgabe liegt die Information

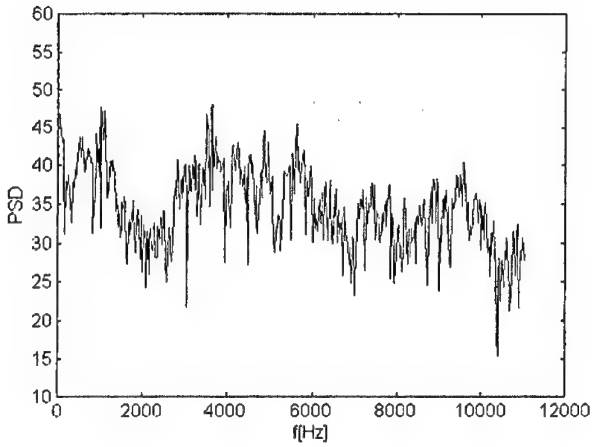

Abbildung 7.30.: Typisches Leistungsdichtespektrum des Frikatives */x/*. Aufgrund der rückverlagerten Engstelle zwischen Zungenrücken und hinterem Gaumen erfährt das rauschartige Anregungssignal (Friktionsgeräusch) durch das Passieren des vorderen Vokaltrakts eine vokalische Färbung.)

bezüglich der Klassenzugehörigkeit in der Abhängigkeit der 22 Datenvektorelemente verborgen.

Die Diskriminanzanalyse ist das am häufigsten angewandte Verfahren der Statistik zur Lösung derartiger Problemstellungen ([Nor93]). Während es bei der 2-Klassen-Diskriminanzanalyse möglich war, mit *einer* Diskriminanzfunktion eine optimale Klassentrennung herbeizuführen, werden bei der 3-Klassen-Diskriminanzanalyse *zwei* Diskriminanzfunktionen benötigt. Sie bestehen aus denjenigen Sätzen von Koeffizienten, die nach Linearkombination mit allen zur Verfügung gestellten Datenvektoren einer Trainings-Stichprobe eine optimale Klassentrennbarkeit sicherstellen bzw. eine Zuordnung mit minimaler Fehlklassifikationsrate gestatten. Die MODIAS-Diskriminanzfunktionen bestehen (wie auch die Datenvektoren) aus jeweils 22 Elementen $B_1 \ldots B_{22}$ bzw. $C_1 \ldots C_{22}$ und einer additiven Konstante B_0 bzw. C_0.

Das Ergebnis der Linearkombinationen beider Diskriminanzfunktionen mit dem Daten-

vektor des j-ten Segmentes $\mathcal{E}_j$ ist das Diskriminanzscorepaar $(D1_j, D2_j)$ mit:

$$D1_j = B_0 + B_1\mathcal{E}_{j_1} + B_2\mathcal{E}_{j_2} + \ldots + B_{22}\mathcal{E}_{j_{22}} = B_0 + \sum_{i=1}^{22} B_i\mathcal{E}_{j_i} \tag{7.28}$$

$$D2_j = C_0 + C_1\mathcal{E}_{j_1} + C_2\mathcal{E}_{j_2} + \ldots + C_{22}\mathcal{E}_{j_{22}} = C_0 + \sum_{i=1}^{22} C_i\mathcal{E}_{j_i} \tag{7.29}$$

Damit wird die Gesamtinformation des ursprünglich 22-elementigen Datenvektors $\mathcal{E}_j$ des j-ten Elements auf ein Wertepaar $(D1_j, D2_j)$ abgebildet; durch Reduktion entsteht aus dem 22-dimensionalen *Datenvektor* ein 2-dimensionaler *Merkmalsvektor*.
Das Diskriminanzscorepaar wird für jedes Signalsegment $j = 1 : J$ bzw. jeden Datenvektor $\mathcal{E}_j$ der gesamten Aufzeichnung ermittelt. Werden alle Merkmalsvektoren der Trainings-Stichprobe über dem nun aufgespannten 2-dimensionalen Merkmalsraum aufgetragen, entsteht ein Scatterplot mit 3 charakteristischen Clustern entsprechend der jeweiligen Klassenzugehörigkeit (Abbildung 7.31).

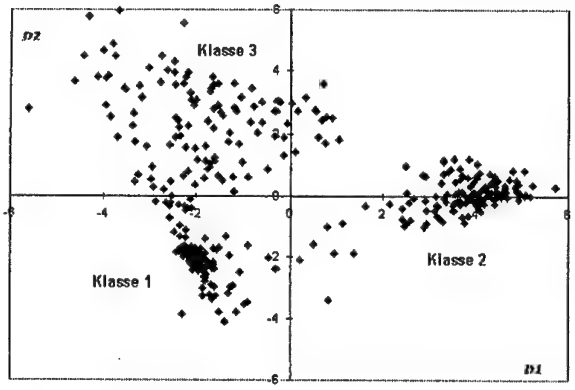

Abbildung 7.31.: Scatterplot im 2-dimensionalen Merkmalsraum mit typischer Clusterbildung entsprechend den 3 betrachteten Klassen

7.3.2.4. Erkennertraining

Das Erkennertraining, also die empirische Optimierung des Mustererkenners zur Lösung der gestellten Klassifikationsaufgabe anhand einer hinreichend großen Stichprobe besteht bei der realisierten Entwurfsstrategie in der Durchführung einer 3-Klassen-Diskriminanzanalyse und damit der Gewinnung zweier geeigneter Diskriminanzfunktionen. Dabei ist für die Klassifikationsaufgabe

- Frikativ /*s*/ gegenüber Frikativ /$\int$/ bzw. Frikativ /*x*/

nur ein einziger Berechnungsgang erforderlich.

Methodischer Ansatz: Die Trainingsphase erfordert eine Handsegmentierung jeder einzelnen Aufzeichnung der Trainings-Stichprobe. Dabei wird über visuelle Prüfung (Oszillogramm) und gleichzeitig perzeptive Kontrolle (Lautsprecher) für jeden Frikativ jeweils ein repräsentatives Signalsegment bzw. dessen zugehöriger Datenvektor extrahiert. Die Information bezüglich der Klassenzugehörigkeit ist dabei bekannt. Die Menge der so entnommenen Datenvektoren der 3 vorkommenden Klassen, jeweils versehen mit der Zusatzinformation der Klasse, bildet die Basis zur Durchführung der Analyse, die im Anschluß mit einem der etablierten Statistik-Programme, hier SPSS, erfolgt. Durch die identische Bildung der Datenvektoren zu Kapitel 7.3.1.3 konnten die hier relevanten Datenvektoren der Frikative direkt übernommen werden, damit entfiel ein weiterer Segmentierungsdurchgang. Die Menge der so entnommenen Datenvektoren der drei Klassen bilden die Basis zur Durchführung der nachfolgenden Analyse.

Am Ende einer 3-Klassen-Diskriminanzanalyse stehen grundsätzlich mehrere Informationen zur Verfügung, das sind

- die beiden gesuchten Diskriminanzfunktionen, die eine optimale Klassentrennbarkeit anhand des Diskriminanzscorepaares $(D1, D2)$ gestatten,
- die drei 2-dimensionalen Verteilungen der Diskriminanzscorepaare, die sich durch Anwendung der beiden Diskriminanzfunktionen auf sämtliche Muster der Trainings-Stichprobe ergeben,

- die Zentroide der drei so entstandenen Verteilungen, also die klassenweise berechneten Mittelwerte der Diskriminanzscorepaare $(D1, D2)$.

Vorläufige Stichprobe: Um für jedes der Einzelwörter jeweils genau einen repräsentativen Datenvektor für den enthaltenen Frikativ zu gewinnen, war hier eine manuelle Segmentierung der Aufzeichnungen erforderlich. Wären die Diskriminanzfunktionen auf Basis der Trainings-Stichprobe zu diesem Zeitpunkt bereits vorhanden, ließe sich die Segmentierung komfortabel in dem grafischen Bewertungseditor (vgl. Abbildung 7.6 auf Seite 205) vornehmen, der den zeitlichen Verlauf der Diskriminanzscores und des zeitsynchronen Oszillogramms darstellt und damit bereits auf die Orte der Frikative innerhalb des Sprachsignales zeigt. In der Phase des Erkennertrainings standen diese Diskriminanzfunktionen aber logischerweise noch nicht zur Verfügung.
Daher wurden zunächst Diskriminanzfunktionen auf Basis einer eigenen Stichprobe berechnet, bestehend aus 1 männlichen und 1 weiblichen Sprecher (siehe Tabelle 7.2 auf Seite 254) mit jeweils 10 Wiederholungen, um zumindest eingeschränkt die grafische Segmentierungsunterstützung nutzen zu können. Die Ergebnisse dieser Voranalyse sind hier weniger von Interesse und werden daher nicht weiter dargestellt. Zusätzlich wurde innerhalb der Editor-Oberfläche eine Spezialroutine implementiert, die

- die Orte der Frikative innerhalb der Aufzeichnung markierte,
- eine leichte Verschiebbarkeit dieser Orte ermöglichte mit paralleler perzeptiver Kontrolle,
- auf Wunsch den Datenvektor an dem gewählten Ort entnahm und automatisch in eine Datenbank portierte.

Auf diese Weise entstanden für jeden der 25 Probanden (vgl. Tabelle 7.3 auf Seite 255) jeweils 12 Datenvektoren pro Frikativklasse (insgesamt waren 36 Wörter zu sprechen, davon je 12 pro Frikativtyp, jedes Wort enthält dabei je einen Frikativ).
Tabelle 7.7 listet sämtliche auf Basis der Trainings-Stichprobe gem. Tabelle 7.3 durchgeführten Diskriminanzanalysen auf mit der jeweils zur Verfügung stehenden Anzahl von Datenvektoren pro Frikativklasse.

Trainings-Stichprobe: Die Gewinnung von Trainingsdaten erfolgte an sprechgesunden Probanden im Rahmen einer Erhebung an der Universität der Bundeswehr München und der Entwicklungsgruppe Klinische Neuropsychologie München-Bogenhausen. Jeder Proband hatte dabei die Aufgabe, sämtliche Einzelwortaufzeichnungen der Patientensitzung des Moduls FRIKATIVARTIKULATION in gewohnter Geschwindigkeit und Stimmlage aufzusprechen. Auf diese Weise entstanden für jeden Probanden Aufzeichnungen von insgesamt 36 Einzelwörtern, davon jeweils 12 für die unterschiedenen Frikativklassen /*s*/, /*ʃ*/ bzw. /*x*/.
Vor jeder Aufzeichnung wurden außerdem folgende Zusatzinformationen erfaßt:

- Alter [Jahre]
- Geschlecht [m/w]
- Raucher [Ja/Nein]

Um eventuelle Verfälschungen der Statistik zu vermeiden, wurde vorsorglich jeder Proband befragt, ob nach seinem eigenen, subjektiven Empfinden eine (wenn auch noch so gering ausgeprägte) Sprechstörung bezogen auf Stimme und/oder Artikulation vorliegt. Dies wurde von keinem der Probanden bejaht.
Die verwendete Trainings-Stichprobe ist identisch zur der bereits in Tabelle 7.3 auf Seite 255 dargestellten Stichprobe. Die im Laufe der Entwicklung zahlreich durchgeführten Diskriminanzanalysen zeigten hierbei, daß durch Aufteilung der Trainings-Stichprobe nach dem Geschlecht wesentlich bessere Klassifikationsergebnisse erreicht werden konnten.

Ergebnisse der Diskriminanzanalysen: Die erforderlichen 3-Klassen-Diskriminanzanalysen gem. Tabelle 7.7 wurden mit dem Statistik-Programmpaket SPSS am Leibniz-Rechenzentrum der Technischen Universität München durchgeführt. Dabei resultierten pro Analyse jeweils die beiden in Tabelle 7.8 und 7.9 aufgelisteten Diskriminanzfunktionen bzw. deren Koeffizienten $B_0 \ldots B_{22}$ bzw. $C_0 \ldots C_{22}$. Die zugehörigen Verteilungsmittelwerte sind in Tabelle 7.10 gesondert zusammengestellt.

7.3.2.5. Klassifikation

Aufgabe der Klassifikation ist es nun, nach Anwendung beider Diskriminanzfunktionen auf ein neues, unbekanntes Muster einer Teststichprobe von dem resultierenden Diskriminanzscorepaar $(D1, D2)$ möglichst sicher auf die Klassenzugehörigkeit zur Klasse k zu schließen.

Die einfachste denkbare Art einer Klassifikationsstrategie wäre die Zuordnung zu der Klasse, deren Zentroid innerhalb der Merkmalsebene am nächsten liegt (Minimum-Distance-Klassifikator gem. Abbildung 7.32).

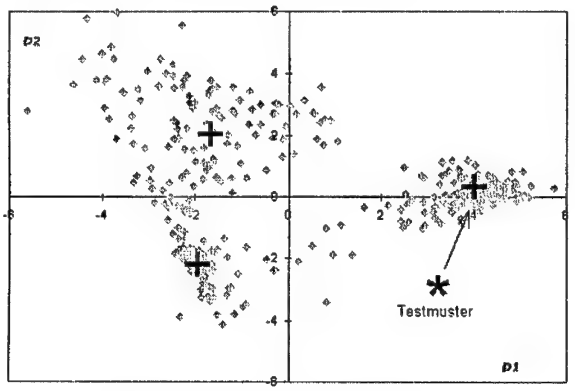

Abbildung 7.32.: Der Minimum-Distance-Klassifikator würde die Zuordnung des Testmusters innerhalb der Merkmalsebene zu der Klasse mit dem nächstgelegenen Klassenzentroid treffen. Dabei bliebe die Verteilung der Merkmale innerhalb der Klassen unberücksichtigt.

Diese Art der Entscheidung berücksichtigt aber nicht die realen klassenspezifischen Verteilungen der Scorepaare um ihren Klassenschwerpunkt. Es ist leicht vorstellbar, daß die Zuordnung zu einer bezogen auf den Klassenzentroid zwar weiter entfernten, aber dafür stärker streuenden Klasse durchaus sinnvoll sein könnte, hier aber unberücksichtigt bliebe.

Da sich jeder Merkmalsvektor aus der gewichteten Summation vieler Einzelelemente zusammensetzt, die ihrerseits jeweils einer spezifischen Streuung unterliegen, erscheint

hier für jede Klasse die Annahme einer 2-dimensional gaußverteilten Wahrscheinlichkeitsfunktion sinnvoll.

Die multivariate Normalverteilung einer Klasse k ist eine Funktion der vektoriellen Zufallsgröße $D = (D1, D2)$; ihre Form wird bestimmt durch den klassenspezifischen Zentroidvektor $m_k = (m1, m2)$ und die klassenspezifische Kovarianzmatrix C_k.
Bei einem 3-Klassenproblem bedeutet dies:

$$p(D|k) = \frac{1}{\sqrt{(2\pi)^3 |C_k|}} \exp^{-\frac{1}{2}(D-m_k)' C_k^{-1} (D-m_k)} \tag{7.30}$$

mit:

$$C_k = (D - m_k)(D - m_k)' \qquad |C_k| = det(C_k) \tag{7.31}$$

Dabei gibt der hochgestellte Strich an, daß es sich hier um einen Zeilenvektor handeln muß. Vektoren ohne diese Markierung sind Spaltenvektoren.
Da die Variable D nur im Exponenten vorkommt, gilt für die Orte gleicher bedingter Wahrscheinlichkeitsdichte (d.h. $p(D|k) = const$) auch:

$$(D - m_k)' C_k^{-1} (D - m_k) = const = q \tag{7.32}$$

Dieser Wert q stellt den *Mahalanobis-Abstand* dar. Damit sind die Orte gleicher Wahrscheinlichkeitsdichte zugleich Orte gleichen Mahalanobis-Abstands. Sie bilden konzentrische Ellipsen bzw. Ellipsoide, deren Hauptachsen in Richtung der Eigenvektoren der zugehörigen Kovarianzmatrix C_k verlaufen. Entlang der Ellipse für Mahalanobis-Abstand 1 ist die bedingte Wahrscheinlichkeit auf $1/\sqrt{e}$ abgefallen (1-Sigma-Ellipse).

MODIAS stellt die aus der Trainings-Stichprobe ableitbaren 1-σ-Ellipsen, 2-σ-Ellipsen sowie die 3 Klassenschwerpunkte $m_1 \ldots m_3$ in der 2-dimensionalen Merkmalsebene dar (Abbildung 7.33). Zusätzlich werden in diese Ebene die resultierenden Scorepaare für die realisierten Frikative des gerade untersuchten Patienten projiziert.
Die Zuordnung des Mustererkenners wird zu der Klasse getroffen, zu der das Muster einen minimalen Mahalanobis-Abstand aufweist.

7.3.3. Auditive Frikativklassifikation durch Expertenurteile

Die auditive Befundung durch geschulte Experten wurde in dem MODIAS-Modul FRIKATIVARTIKULATION derart realisiert, daß jedes aufgezeichnete Einzelwort perzeptiv

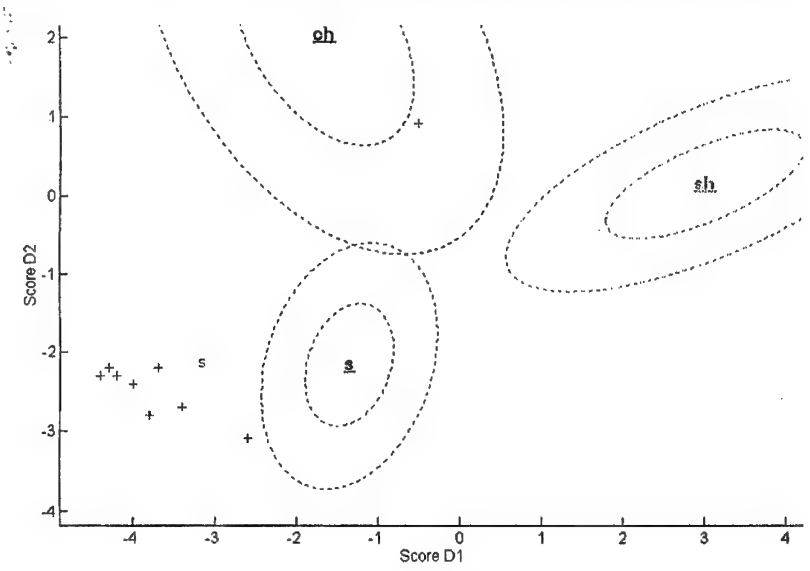

Abbildung 7.33.: MODIAS-Darstellung der Merkmalsebene mit 1-σ und 2-σ-Ellipsen der Trainings-Stichprobe. Zusätzlich sind die Muster projiziert, die sich aus der Frikativrealisierung des gerade betrachteten Patienten ergeben (hier für den Frikativ /s/). Alle (Test-)Muster liegen hier außerhalb der 2-σ-Ellipse der Normgruppe.

dargeboten wird. Die Wiedergabe jedes Wortes kann dabei nur einmal gestartet werden; unmittelbar danach entscheidet sich der Untersucher je nach perzeptiver Wahrnehmung für einen der 3 Frikativtypen. Die Reihenfolge der Wiedergabe bezogen auf Wort- bzw. Frikativtypen erfolgt dabei randomisiert.

Bark-Band Nr.	Untere Grenzfrequenz[Hz]	Obere Grenzfrequenz [Hz]
1	0	101
2	101	204
3	204	309
4	309	417
5	417	531
6	531	651
7	651	781
8	781	922
9	922	1079
10	1079	1255
11	1255	1457
12	1457	1692
13	1692	1971
14	1971	2309
15	2309	2726
16	2726	3247
17	3247	3903
18	3903	4729
19	4729	5758
20	5758	7030
21	7030	8605
22	8605	10583

Tabelle 7.1.: Obere und untere Frequenzgrenzen der 22 gehörrichtigen Bark-Bänder.

Gruppe	Geschlecht	Anzahl	Alter	Raucher	Wiederholungen
1	männlich	1	29	Nein	10
2	weiblich	1	28	Nein	10

Tabelle 7.2.: Zusammensetzung der vor der eigentlichen Trainingsphase gewonnenen Stichprobe zur Bestimmung vorläufiger Diskriminanzfunktionen.

Gruppe	Geschlecht	Anzahl	Alter [Median (Min/Max)]	davon Raucher
1	männlich	14	38 (24/73)	1
2	weiblich	11	30 (20/62)	1

Tabelle 7.3.: Zusammensetzung der Trainings-Stichprobe sprechgesunder Probanden. Für die Gruppe der männlichen und der weiblichen Probanden wurde jeweils eine eigene Diskriminanzanalyse berechnet, um Einflüsse der Sprechstimmlage zu eliminieren.

Analyse	Gruppe	Aufgabe	Datenvektoren Frikativ	Datenvektoren Vokal
1	1 (männlich)	/s/ ↔ /a/	168	168
2	2 (weiblich)	/s/ ↔ /a/	132	132
3	1 (männlich)	/ʃ/ ↔ /a/	168	168
4	2 (weiblich)	/ʃ/ ↔ /a/	132	132
5	1 (männlich)	/x/ ↔ /a/	168	168
6	2 (weiblich)	/x/ ↔ /a/	132	132

Tabelle 7.4.: Übersicht über die durchgeführten 2-Klassen-Diskriminanzanalysen und die jeweilige Anzahl von Trainings-Datenvektoren im Rahmen des Erkennertrainings an 25 sprechgesunden Probanden.

Koeffizient	Analyse 1	Analyse 2	Analyse 3	Analyse 4	Analyse 5	Analyse 6
B_0	5.90	6.91	3.83	4.37	4.06	4.27
B_1	0	0	0	0	0	0
B_2	0	0	0	0	0	0
B_3	-0.27	-0.16	-0.35	-0.12	-0.17	-0.08
B_4	-0.35	-0.62	0	-0.67	0	-0.37
B_5	-0.10	-0.17	-0.15	-0.24	0	-0.16
B_6	-0.11	-0.14	-0.10	-0.17	-0.09	0
B_7	-0.14	-0.15	-0.13	-0.11	-0.13	-0.09
B_8	-0.13	-0.14	-0.12	-0.11	-0.09	-0.14
B_9	-0.12	-0.15	-0.10	-0.12	-0.06	-0.10
B_{10}	-0.11	-0.15	-0.12	-0.15	-0.06	-0.11
B_{11}	-0.13	-0.13	-0.12	-0.13	0	-0.02
B_{12}	-0.13	-0.15	-0.21	-0.12	-0.03	-0.04
B_{13}	0	-0.12	0.06	-0.12	0	0
B_{14}	-0.10	0	0.03	0	0	0
B_{15}	-0.09	0	0.02	0.11	0	0
B_{16}	-0.10	0	0.04	0.04	0	0
B_{17}	0	0	0.03	0.03	0	0
B_{18}	0.05	0	0.07	0.06	0.12	0.03
B_{19}	0	0	0	0	0	0
B_{20}	0	0	0	0	0	0
B_{21}	0	0	0	0	0	0
B_{22}	0	0	0	0	0	0

Tabelle 7.5.: Diskriminanzfunktionen der durchgeführten 2-Klassen-Diskriminanzanalysen im Rahmen des Erkennertrainings. Die Analysen 1-6 sind in Tabelle 7.4 spezifiziert.

Analyse	Verteilungsmittelwert Klasse Frikativ	Verteilungsmittelwert Klasse Vokal
1	5.95	-5.95
2	6.70	-6.70
3	6.63	-6.63
4	7.14	-7.14
5	3.07	-3.07
6	3.31	-3.31

Tabelle 7.6.: Verteilungsmittelwerte der Diskriminanzscores für die durchgeführten 2-Klassen-Diskriminanzanalysen im Rahmen des Erkennertrainings. Sie entstehen, wenn sämtliche Datenvektoren der Klasse Frikativ oder der Klasse Vokal (vgl. Tabelle 7.4) mit der jeweiligen Diskriminanzfunktion aus Tabelle 7.5 linear kombiniert werden und über die so erhaltenen Diskriminanzscores gemittelt wird.

Analyse	Gruppe	Datenvekt. /s/	Datenvekt. /ʃ/	Datenvekt. /x/
1	1 (männlich)	168	168	168
2	2 (weiblich)	132	132	132

Tabelle 7.7.: Übersicht über die durchgeführten 3-Klassen-Diskriminanzanalysen im Rahmen des Erkennertrainings.

Koeffizient	Analyse 1	Analyse 2
B_0	-1.46	-2.29
B_1	0	0
B_2	0	0
B_3	-0.1	-0.10
B_4	0	0
B_5	0	0
B_6	0.03	0
B_7	-0.33	0.12
B_8	0	0
B_9	-0.07	-0.13
B_{10}	-0.04	-0.17
B_{11}	-0.01	-0.01
B_{12}	-0.04	-0.05
B_{13}	-0.02	-0.05
B_{14}	0.08	0.08
B_{15}	0.06	0
B_{16}	0.10	0.08
B_{17}	0.05	0.07
B_{18}	0.04	0.04
B_{19}	-0.01	0.1
B_{20}	-0.01	0
B_{21}	0	0
B_{22}	0	0

Tabelle 7.8.: Erste Diskriminanzfunktion $B_0 \ldots B_{22}$ der durchgeführten 3-Klassen-Diskriminanzanalysen im Rahmen des Erkennertrainings.

Koeffizient	Analyse 1	Analyse 2
C_0	-0.96	-1.74
C_1	0	0
C_2	0	0
C_3	0.13	0.07
C_4	0	0
C_5	0	0
C_6	0.2	0
C_7	0.13	0.56
C_8	0	0
C_9	0.09	0.25
C_{10}	0.07	0.16
C_{11}	0.04	0.04
C_{12}	0.06	0.08
C_{13}	0.05	0.08
C_{14}	0.04	0.02
C_{15}	0.01	0
C_{16}	0.02	0.02
C_{17}	0.03	0.01
C_{18}	0	0.04
C_{19}	-0.02	0.03
C_{20}	-0.04	-0.05
C_{21}	0	0
C_{22}	0	0

Tabelle 7.9.: Zweite Diskriminanzfunktion $C_0 \ldots C_{22}$ der durchgeführten 3-Klassen-Diskriminanzanalysen im Rahmen des Erkennertrainings.

Analyse	Klassenzentroid /s/	Klassenzentroid /ʃ/	Klassenzentroid /x/
1	-1.36 , -2.16	3.02 , 0.14	-1.66 , 2.02
2	-1.86 , -2.16	3.77 , 0.02	-1.91 , 2.14

Tabelle 7.10.: Diskriminanzscore-Verteilungsmittelwerte (Zentroide) der durchgeführten 3-Klassen-Diskriminanzanalysen im Rahmen des Erkennertrainings. Sie entstehen, wenn sämtliche Datenvektoren der 3 Frikativklassen (vgl. Tabelle 7.7) nacheinander mit den jeweiligen Diskriminanzfunktionen aus Tabelle 7.8 bzw. 7.9 linear kombiniert werden und über die so erhaltenen Diskriminanzscorepaare gemittelt wird.

Zusammenfassung

Ungefähr 2 Prozent der Bevölkerung der westlichen Welt leiden an den Nachwirkungen einer Hirnschädigung. In vielen Fällen sind mit einer solchen Schädigung neurogene Sprechstörungen verbunden, die alle funktionellen Strukturen des Sprechvorganges wie Sprechatmung, Stimme und Artikulation betreffen können.
Eine eingehende Analyse der Arbeitsmethoden in der klinischen Diagnostik neurogener Sprechstörungen hat gezeigt, daß hier noch vorwiegend auditive Verfahren zum Einsatz kommen, d.h., ein geschulter Hörer beurteilt einen Patienten nach seinem persönlichen Höreindruck. Solche Hörersitzungen sind zeitintensiv und subjektiv. Akustisch-phonetische Verfahren, also der unterstützende Einsatz technischer Systeme, tragen hier zu einer Objektivierung des klinischen Bildes bei. Obwohl hierzu im Prinzip bereits die notwendigen Methoden und Verfahren verfügbar sind und eine breite klinische Nutzung befürwortet und auch angestrebt wird, finden sich bis heute nicht einmal die einfachsten akustischen Parameter routinemäßig in klinischen Befunden und Berichten.
Die Frage nach der Ursache dieser Situation erbrachte folgende Erkenntnisse:

- Technische Diagnostik-Systeme sind oft zu kostenintensiv, gemessen am vergleichsweise geringen Budget neuropsychologischer Einrichtungen.
- Sie enthalten häufig zu viele Freiheitsgrade und liefern eine schwer überschaubare Anzahl von Parametern mit teilweise geringer diagnostischer Relevanz. Zur Erzeugung der Parameter sind häufig zeitaufwendige Vorverarbeitungen erforderlich.
- Sie enthalten keine standardisierten und vorgefertigten Untersuchungsprotokolle.
- Die Ergebnisse sind häufig nicht transparent genug und für einen wenig technisch versierten Anwender nur schwer interpretierbar.

Zusammenfassung

Hier klafft eine deutliche Lücke zwischen dem technisch Machbaren und dem diagnostisch Sinnvollen. Aus dieser konkreten Problemstellung heraus entstand im Rahmen der vorliegenden Arbeit das **Modulare Diagnostik-System für Sprechstörungen (MODIAS)**. Dieses System ist das Resultat einer über 5jährigen direkten und fruchtbaren Kooperation des Instituts für Nachrichtentechnik der Universität der Bundeswehr München mit der Entwicklungsgruppe klinische Neuropsychologie des Städtischen Krankenhauses München-Bogenhausen. Durch das interdisziplinäre Zusammenwirken neuropsychologischer Expertise mit ingenieurwissenschaftlicher Methodik konnten Synergieeffekte gewinnbringend genutzt werden, um die komplexe Aufgabenstellung mit der Vielzahl involvierter Fachdisziplinen wie der Neuropsychologie, Neurolinguistik und Phonetik auf der Anwenderseite sowie der Elektrotechnik und Informatik auf der Entwicklerseite entstehen zu lassen. Die technische Lösung der gestellten Aufgaben streifte dabei weite Gebiete der **Digitalen Signalverarbeitung**, **Biosignalverarbeitung** und **Psychoakustik** bis hin zur **Mustererkennung** und der **Künstlichen Intelligenz**.

Die Erfahrungen mit dem System **TUS** und dem computerunterstützten Experimentiersystem PhonX - beide Ergebnisse früherer Kooperationen der beiden genannten Institutionen - gaben dabei wertvolle Impulse für den gesamten Entwicklungsprozeß. Während das TUS-System zur automatisierten Prüfung des Teilaspekts *Verständlichkeit* nach dem sogenannten *Müncher Verständlichkeitsprofil* dient ([Ahr94], siehe ergänzend auch [Fin90]), deckt das System PhonX Teilaspekte der Diagnose und Therapie von zentralen Störungen der Produktion (auf der höheren Ebene der Sprechplanung) und des Verstehens gesprochener Sprache ab ([Tei95]). Beide Systeme sind heute im Einsatz, daher wurden schon zu Beginn der hier vorgestellten Entwicklung Datenschnittstellen geschaffen und damit die Kompatibilität hergestellt.
Mit **MODIAS** ist es jetzt gelungen, dem Diagnostiker darüberhinaus erstmals ein **umfassendes Instrument zur Standarddiagnostik** neurogener Sprechstörungen anhand akustischer Analysen des Sprachsignals und **zur Befunderstellung anhand akustischer Parameter** an die Hand zu geben. Hierbei finden neueste Methoden der Digitalen Signalverarbeitung Anwendung. Die folgenden Leistungsmerkmale qualifizieren MODIAS in herausragender Weise für einen **klinischen Einsatz** in der Sprechstörungsdiagnostik:

- MODIAS liefert nur klinisch relevante Parameter, die geeignet sind, valide Aussagen

über die wichtigsten Aspekte wie Artikulation, Stimme und Prosodie zu treffen.

- MODIAS integriert alle wichtigen Aspekte einer klinischen Standarddiagnostik von der Präsentation des Stimulusmaterials über die Sprachaufzeichnung mit anschließender Berechnung akustischer Parameter bis hin zur Ergebnispräsentation.
- Alle akustischen Parameter sind technisch bezüglich Stabilität und Exaktheit optimiert. Sämtliche Berechnungen erfordern dabei eine nur minimale Benutzerinteraktion.
- Ergebnisse werden klar und übersichtlich präsentiert; sie sind leicht interpretierbar und direkt in klinische Berichte und Befunde portierbar.

MODIAS ist ein PC-gestütztes System, aufbauend auf dem mathematischen Interpreter MATLAB. Es besteht aus **4 unabhängigen Modulen**.

Im Modul **Stimme und Vokalartikulation** wird der Patient aufgefordert, die vier Vokale */i/,/y/,/u/* und */a/* über mehrere Sekunden in normaler Stimmlage und Lautstärke anzuhalten. MODIAS stellt dem Diagnostiker akustische Parameter zur Verfügung, die geeignete Maße für die Stellung der relevanten Artikulatoren wie Lippen, Zunge, Kiefer und Gaumensegel darstellen und damit Aufschluß über die Fähigkeit des Patienten zu einer kontrastreichen Vokalartikulation geben. Die Analyse der Stimmbandschwingung erlaubt zusätzlich die Beurteilung von Qualität und Stabilität der Stimmgebung. Die technische Realisierung des Anteils *Stimme* erforderte die Entwicklung eines Pitch-Tracking-Verfahrens, das im Zeitbereich jede einzelne Stimmbandschwingung bezüglich ihrer Periodendauer bzw. Frequenz vermißt. Die Stabilität der so erhaltenen Tonhöhenverläufe wird nach einer neuen Methode analysiert, die eine differenzierte Betrachtung nach verschiedenen Arten von Instabilität (Pitch-Trend, Pitch-Variabilität und Jitter) gestattet. Zur Beurteilung der Stimmqualität wurde unter anderem der erst vor kurzem vorgestellte CPPS-Parameter (Cepstral Peak Prominence Smoothed) als Behauchtheits-Prädiktor implementiert, der hier zur Beurteilung von intermittierenden Störungen erstmals auch in seinem zeitlichen Verlauf verfolgt werden kann. Die Validität der beiden Stimmqualitätsmaße für Rauhigkeit (Jitter) und Behauchtheit (CPPS) wurde anhand synthetischer Stimmen überprüft. Hierzu war die Entwicklung eines Formant-Vocoders und die Simulation

der Stimmbandschwingung mit beiden Stimmstörungsmerkmalen erforderlich. Die technische Realisierung des Anteils *Vokalartikulation* führte zur Modellierung des menschlichen Vokaltrakts mit seinen Sprechwerkzeugen durch ein Allpol-Filter (Autoregressives Filtermodell), dessen konjugiert komplexe Polpaare der korrespondierenden Systemfunktion über das Verfahren der Linearen Prädiktion nach der Autokorrelationsmethode unter Zuhilfenahme der Levinson-Durbin-Rekursion bestimmt werden. Bedingt durch die Zeitvarianz des Sprachsignals werden diese Polpaare (Modellparameter) 100 mal pro Sekunde neu berechnet. Der Rückschluß von der Lage der Polpaare in der z-Ebene auf die Frequenzlage der gesuchten charakteristischen Überhöhungen in den korrespondierenden Fourier-Übertragungsfunktionen (*Formanten*) erfolgt hier nach einem neuen nichtlinearen Berechnungsverfahren. Dabei konnte gezeigt werden, daß das bisher bekannte lineare Berechnungsverfahren in ungünstigen Fällen zu inakzeptablen Fehlern in der Formantschätzung führen kann. Zusätzlich konnten dadurch kritische Pol-Lagen der z-Ebene angegeben werden, die zu extremen Varianzen der abgeleiteten Formantfrequenzen führen müssen (Ursache von *Spurious Peaks* und *Drop-Outs*), und überkritische Bereiche, in denen auch konjugiert komplexe Polpaare nicht resonanzfähig sind (*Spectral Shaping Poles*). MODIAS detektiert und unterdrückt kritische und überkritische Polpaare und liefert dadurch besonders robuste Formantverläufe. Da bei sprechgestörten Patienen jegliche Normwerte für Formantlagen fehlen, kann der zusätzlich entwickelte Formant-Tracking-Algorithmus im Falle unplausibler Daten manuell beeinflußt werden.

Das Modul **Satzproduktion** widmet sich den Veränderungen von Tempo, Rhythmus und Redefluß. Hier wurde eine PC-gesteuerte Methode zur akustischen Analyse des Zeitmusters der Satzproduktion entwickelt. Der Patient hat die Aufgabe, ausgewählte Testsätze nachzusprechen, die sich bei festem Trägersatz in nur einem Zielwort variierter Komplexität unterscheiden (z.B. Ute kann die *Trompete* bekommen) oder bei denen zusätzlich auch der Trägersatz variiert wurde. Das Verfahren erlaubt gleichzeitig eine auditive Analyse nach phonetischen und phonematischen Fehlern. Es eignet sich damit zur Differenzierung unterschiedlicher Störungsmuster und zur Analyse der Faktoren, die das Auftreten von Unflüssigkeiten beeinflussen. Zur technischen Realisierung wurde hier das von Ruske und Beham vorgeschlagene vereinfachte Funktionsmodell der Lautheit implementiert. Es berücksichtigt die Vorgänge im menschlichen Gehör durch Simulation der Erregungsverteilung entlang der Basilarmembran in Abhängigkeit des anliegenden

Schalldruckverlaufs. Durch den Übergang von physikalischen Meßgrößen auf psychoakustische Empfindungsgrößen wird dem anschließenden Silbendetektionsalgorithmus die gleiche Grundlage zur Verfügung gestellt, die auch der Mensch zur Auflösung des Gesprochenen in einzelne Silben heranzieht. Zur präziseren Detektion von Silbengrenzen wurde zusätzlich ein Verfahren entwickelt, mit dem die Rausch- und Brummstörungen innerhalb des Aufzeichnungssystems PC/Soundkarte nach Methoden der statistischen Signaltheorie analysiert und entsprechend berücksichtigt werden können.

Das Modul **Schnelle Silbenwiederholungen** gestattet die akustische Analyse der Fähigkeit des Patienten zu schnellen Silbenwiederholungen (Diadochokinese). Hierbei sind Silbenfolgen, wie z.B. *babababa* oder *danadana*, so schnell und so lange wie möglich zu realisieren. Von Interesse sind hier besonders Tempowechsel sowie dysrhythmische Tonhöhe- und Lautstärkeschwankungen. Technisch konnten hier Methoden des Moduls Satzproduktion übernommen werden, erweitert um neue akustische Parameter, die die Regularität der Silbenproduktion und deren Störungen differenziert beschreiben.

Das Modul **Frikativartikulation** widmet sich der akustischen Analyse von stimmlosen Zischlauten (Frikative). Der Patient spricht ausgewählte Einzelwort-Tripel wie z.B. *lassen, Laschen, lachen* nach. Von Interesse ist hier einerseits die Fähigkeit zur Kontrastierung der eingebetteten Frikative gegenüber dem jeweils vorangehenden Vokal */a/* (Diskrimination) sowie zur Unterscheidung der 3 verschiedenen Frikative (Differenzierung). Technisch war hierzu der Entwurf und die Realisierung von insgesamt 8 verschiedenen Mustererkennern erforderlich, die jeweils an einer hinreichend großen Stichprobe sprechgesunder Probanden trainiert wurden. Ein neues Verfahren der Merkmalsextraktion, Merkmalsreduktion und der Normierung ermöglicht es hierbei, alle relevanten Laute aussteuerungs- und intensitätsunabhängig (d.h. unabhängig vom Aufzeichnungssystem und von der Stimme) ausschließlich anhand ihrer spektralen Verteilungen zu klassifizieren. Für die Komponente *Diskrimination* wurde ein Verfahren entwickelt, mit dem die Qualität der Artikulation anhand der Zugehörigkeitswahrscheinlichkeit zu einer der beiden Lautklassen beurteilt werden kann (Bayes-Klassifikator). Hierzu wurden die klassenspezifischen Merkmalsverteilungen durch 1-dimensional gaußverteilte Wahrscheinlichkeitsdichtefunktionen approximiert und die statistischen Rückschlußwahrscheinlichkeiten auf die jeweilige Klasse herangezogen. Für die Komponente *Differenzierung* wurde ein Mustererkenner realisiert, dessen Fehlklassifikationsrate

als akustischer Prädiktor zur Beurteilung der Frikativunterscheidung dient. Als Klassifikationsstrategie wurde hier eine Mahalanobis-Abstandsklassifikation gewählt, die sich unter anderem einer Approximation der klassenspezifischen Merkmalsverteilungen durch 2-dimensional gaußverteilte Wahrscheinlichkeitsdichtefunktionen bedient. Zusätzlich wurde eine auditive Entscheidungsinstanz implementiert, um die auf der Basis akustischer Parameter getroffene Entscheidung des Mustererkenners mit der anhand des Höreindrucks getroffenen Entscheidung des Menschen vergleichen zu können.

Die Entwicklung des MODIAS-Systems war von Anfang an geprägt von einem **kontinuierlichen Dialog** mit dem klinischen Anwender (Mediziner bzw. Therapeuten), der zu einem iterativen Optimierungsprozeß führte. Die **klinischen Erfahrungen** resultierten dabei in schrittweisen Anpassungen von Algorithmen und Oberflächen; in Einzelfällen mußten Teilentwicklungen auch abgebrochen und mit völlig neuen Lösungsansätzen fortgesetzt werden. Dabei wurden bis heute mehr als 500 Menschen untersucht.

MODIAS wurde bereits vor der Veröffentlichung dieser Arbeit auf Fachkongressen in Deutschland, England, den Niederlanden, Österreich und Italien mit großer Resonanz vorgestellt. Dabei konnten zahlreiche Kontakte für Kooperationsprojekte geknüpft werden, wie z.B. zur **Universitätsklinik Heidelberg** im Rahmen eines experimentellen Hirnforschungsprojektes. Hier soll MODIAS für eine Patientengruppe mit schwerem Parkinsonsyndrom eingesetzt werden, um die durch elektrische Stimulation der Stammganglien bewirkte Veränderung des Sprechstörungsgrades zu dokumentieren. Die Forscher erhoffen sich durch den Einsatz des MODIAS-Systems als **Instrument der experimentellen Forschung** wertvolle Erkenntnisse über neue Behandlungsmethoden dieser Erkrankung und auch über die Funktionalität des menschlichen Gehirns.

Eine weitere Kooperation mit der **Universität Essen** wird sich mit der postoperativen sprachlichen Untersuchung von Kindern mit Tumoren in der hinteren Schädelgrube (Kleinhirn) befassen.

Besonders erwähnenswert ist die langjährige Kooperation mit der Polyklinik für Mund-, Kiefer- und Gesichtschirurgie des **Klinikums Rechts der Isar**. Der dortige Einsatz des MODIAS-Systems erbrachte auch wertvolle Erfahrungen mit Patienten

einer rein organisch, also nicht neurogen bedingten Sprechstörung. Im Rahmen der bisher weltweit umfassendsten akustischen Reihenuntersuchung an 154 Patienten mit Lippen-Kiefer-Gaumenspalten konnten damit Aussagen bezüglich der Prävalenz von Stimmstörungen dieser Patientengruppe geliefert und auch wertvolle Hinweise bezüglich der Effektivität neuer Operationstechniken gewonnen werden.

Die Entwicklungsgruppe klinische Neuropsychologie des Städtischen Krankenhauses **München-Bogenhausen** konnte durch den Einsatz des MODIAS-Systems im Rahmen mehrerer Projekte grundlegende Erkenntnisse über die Leistungen bei Satzproduktion und Diadochokinese für Patienten verschiedener neurogener Sprechstörungssyndrome gewinnen (Syndromvergleiche). Zusätzlich liegen Ergebnisse von Langzeit-Studien über die Therapieverläufe schwer sprechgestörter Patienten vor (Einzelfallstudien). MODIAS hat sich hier bereits als **fester Bestandteil der klinischen Diagnostik und Therapie etabliert**, und es liefert zusätzlich einen Beitrag zu aktuellen Fragestellungen der linguistischen, psychologischen und medizinischen Forschung.

Die geschilderten klinischen Anwendungen bildeten die Basis für sprachwissenschaftliche Magister- und Doktorarbeiten mit zahlreich publizierten Einzelergebnissen und -beiträgen.

Für die Zukunft ist - neben diversen klinischen Anwendungen - eine Migration auf die neueste MATLAB-Version geplant. Im Anschluß soll die erst seit diesem Jahr mögliche Kompilierung des MATLAB-basierten MODIAS-Source-Codes zu einer selbständig lauffähigen Applikation für jede beliebige Rechnerplattform folgen, um die Systemkosten nochmals deutlich zu reduzieren und damit die klinische Nutzung weiter voranzutreiben.

Summary

Approximately 2 per cent of the population in the Western World suffer from the longterm-effects of a brain damage. These disorders are frequently associated with speech impairments, which can affect all functional structures of the speech process, such as respiration, voice and articulation.
A detailed analysis of the methods used in speech diagnostics has shown a prevalence of auditory methods where an experienced listener is faced with the problem of assessing and classifying the disorder and of estimating its severity reliably. Auditory methods are time-consuming and subjective. Acoustic analysis techniques, on the other hand, can provide clinicians with appropriate tools to reduce their diagnostic workload and to objectify their treatment effects. Although in principle we already dispose of the required methods and although their wide clinical use is highly recommended, not even the most straightforward parameters are regularly included as a standard in patient reports so far.
A couple of reasons can be named to explain the technological lag in routine clinical diagnostics:

- Systems are often too expensive, compared to the comparatively low budget of neuropsychological institutions.
- Most systems contain too many degrees of freedom and they partly produce data with questionable diagnostic relevance.
- To obtain stable results, most systems require time-consuming preparatory work (e.g. for speech wave segmentations).
- Standard protocols for routine diagnostic examinations are lacking.

- Presentation of results is often not sufficiently transparent and the interpretation of the obtained parameters may be quite difficult for clinical users.

Starting out from the fact that there is a clear gap between technologically feasible and clinically available methods the present thesis aimed at developping a **Modular Diagnostic System for Speech Disorders (MODIAS)**. The system presented here is the result of more than five years of direct and fruitful cooperation between the Institute for Communication Engineering of the Federal Armed Forces University, Munich, and the Clinical Neuropsychology Research Group (EKN) of the City Hospital München-Bogenhausen. The interdisciplinary cooperation between experts in clinical neuropsychology and communication engineers produced synergy effects which could be used profitably in the formulation of the task and in the realization of the instrument. The project interfaced with several special fields, such as Neuropsychology, Neurolinguistics, and Phonetics on the side of the user as well as Electrical Engineering and Information Sciences on the side of the system designer. From the technical point of view, the implementation of the MODIAS system presented here touched large parts of the following scientific fields: **Digital Signal Processing**, **Psychoacoustics**, **Pattern Recognition**, and **Artificial Intelligence**.

Experiences made with earlier offsprings from the cooperation mentioned above, such as the TUS system or the computer-aided experimental system PhonX, gave valuable impulses for the whole process of development. While the TUS system covered the aspect of *intelligibility* measurement (according to the *Munich Intelligibility Profile*), the PhonX system tapped into more specific aspects of diagnostics and treatment of neurogenic disorders of speech production (on the higher level of speech planning) and of speech comprehension. Since both systems are in use today, data interfaces had to be developed at the beginning of the design process, in order to establish compatibility.
The MODIAS system provides speech pathologists with a **powerful and extensive instrument for routine diagnostics** of neurogenic speech disorders. The system is based on acoustic measures obtained from the microphone signal and uses current methods of Digital Signal Processing. The following performance profile qualifies MODIAS especially for diagnostic use in **a clinical environment**:

- MODIAS provides only clinically relevant speech parameters focused on the most

important features of articulation, voice, and prosody.

- MODIAS integrates all important aspects of a clinical examination based on acoustic measures, including stimulus presentation, speech signal recording, computation of parameters, and presentation of results.
- All acoustic parameters are technically optimized with regard to stability and precision. Data processing requires only minimal intervention by the user.
- Results are presented clearly and comprehensively; they can easily be interpreted and directly be transferred to patient reports.

MODIAS is based on a PC platform with a commercial sound-card and uses MATLAB as a programming environment. It consists of **four independent modules**.

In the **Voice and Vowel Articulation** module patients are required to sustain each of the four cardinal vowels */i/*, */y/*, */u/* and */a/* at a comfortable pitch and loudness level over several seconds. Parametric analysis is based on measures describing the position of predominantly the lips and the tongue thereby giving information about the patient's ability of articulating the different vowels in a sufficiently contrastive manner. Time-domain and cepstral analyses of the oscillogram additionally permit to judge voice quality and voice stability. The technical realization of the submodule *Voice* required the design of a time-based pitch-tracking-algorithm that detects fundamental periods cycle-by-cycle. The stability of the resulting pitch trajectories is analyzed by using a new method which allows for a detailed examination of different kinds of relevant types of instabilities (pitch-trend, pitch-variability and jitter). As a measure of breathy voice quality, the CPPS parameter (Cepstral Peak Prominence Smoothed), introduced only recently, was implemented. In particular, CPPS-over-time-trajectories can be analyzed in order to detect intermittent perturbations, a feature that has not been used before. The validity of the two measures of roughness (jitter) and breathiness (CPPS) was established by a set of synthetic voice samples. This method required the development of a formant-vocoder and the simulation of vocal cord vibrations with different roughness and breathiness ratios. The technical realization of the submodule *Articulation* was based on modelling the vocal tract by means of an autoregressive filter design. In this model, all conjugate complex pole-pairs of the system functions are computed by

the autocorrelation method followed by a Levinson-Durbin-Recursion. Because of the time-variant nature of the speech signal, all pole-pairs (model parameters) are calculated 100 times per second. The correlation between the pole-pair location in the complex z-plane and the excessive maximum of the corresponding Fourier-transform (*formants*) follows a new, non-linear calculation algorithm, since the conventional linear method showed inacceptable errors in formant analysis in unfavourable cases. In addition, this method allows for an identification of critical areas within the z-plane causing extreme variances of the obtained formant frequencies (*spurious peaks* and *dropouts*), and of overcritical areas where even conjugate complex pole pairs lack of any detectable resonance at all (*spectral-shaping-poles*). MODIAS detects and supresses these critical and overcritical pole-pairs and thereby provides stable formant trajectories even in cases of poor maintenance of the speakers laryngeal setting over time. Since unplausible formant frequency regions cannot be defined for patients with speech disorders, the formant-tracking-algorithm provides interactive tracking facilities for the handling of questionable data.

Alterations in speech rate, rhythm, and fluency can be examined by the **Sentence Production** module. For this purpose a method has been developed for the assessment of the temporal structure of an utterance. The patient is required to speak a standard set of test sentences which are composed of a constant carrier phrase with embedded target words of varying complexity (Ute kann die *Trompete* bekommen). In an alternative version, the carrier phrase is varied, in order to avoid learning effects in close-meshed treatment control applications. A simplified loudness model developed by Ruske and Beham is used to automatically detect syllable boundaries in the signal. By this algorithm, the nature of human hearing is reproduced by a simulation of the distributed excitation along the basilar membrane as a function of the corresponding acoustic waveform. By changing from physical measurements into psychoacoustic dimensions the syllable detection algorithm uses a similar speech segmentation strategy as human listeners do. The detection of syllable boundaries is further improved by a procedure which compensates for noise disturbances caused by the PC-/soundcard system on the basis of statistical signal theory methods. The results of automatic syllable segmentation can be inspected and corrected interactively, using both visual and auditory information. The temporal information on the syllabic grid of a sentence can be supplemented by auditory-based information on segmental errors. Syllables can

be marked for such errors, yielding a combined time-accuracy pattern of articulation which may be useful in the examination of sentence production disorders in aphasia and apraxia of speech.

The **Diadochokinesis** module allows for a PC-based examination of rapid syllable repetition tasks. Syllable repetition chains such as */bababa.../* or */danadana.../* are to be realized as fast and as long as possible. Evaluation is focused on rate and on alterations or dysrhythmic variations of syllable duration and loudness. A syllable detection algorithm is applied which is based on the same parameters as in the Sentence Production module, but extended by specific measures describing irregularities in syllable production. This module applies as a standard instrument for the sensitive assessment of dysarthria severita and, as it turned out by now, for the differential diagnosis of dysarthria and apraxia of speech.

The module **Fricative Articulation** is a tool for the acoustic analysis of voiceless fricatives. The patient is required to produce a fixed set of phonetic minimal triples, such as *lassen, Laschen, lachen.* The examination protocol is focussed on the assessment of the faculty of contrasting the fricatives */s/*, */sh/*, and */x/* versus preceding vowels as well as of the faculty of realizing the three different fricatives in a sufficiently distinctive manner. For this purpose, eight different pattern recognition systems were designed and each of them was trained on a sufficient number of normal speakers. New methods for feature-extraction, feature-selection and feature-normalisation allow for a classification of the selected speech sounds, exclusively from their spectral characteristics. Irrelevant factors such as absolute loudness or system parameters like predefined gain levels of the recording system are controlled. For the discrimination between vowels and fricatives a procedure has been developed to measure probabilities of class membership (Bayes-classifier) as a predictor for articulatory discrimination capability. For this purpose, class-specific feature distributions were approximated by 1-dimensional Gaussian densities in order to evaluate a-posteriori-probabilities for each class. For the between-fricative-discrimination component a special pattern recognition system was realized which uses misclassification rates to predict imprecise fricative articulation. Here a Mahalanobis-distance-classifier was implemented to approximate class-specific feature distributions by 2-dimensional Gaussian densities. Additionally, auditory based ratings have been implemented in order to compare acoustic-based classifica-

tions of the pattern recognition system with auditory-based decisions of a trained listener.

From the beginning, the development of the MODIAS system has been steered by a **continuous dialogue** between system developer and clinical users, which has led to an iterative optimisation process. The overall design of the system was inspired by **clinical needs** and by previous **clinical and methodological research**. This was based on a close collaboration with the Clinical Neuropsychology Research Group (EKN) of the City Hospital München-Bogenhausen. Further, clinical experience contributed to a step-by-step adjustment of algorithms and user interfaces. In particular cases, partial developments had to be stopped and resumed by a entirely new approach. So far, more than 500 subjects have been examined.

Before the publication of this work, MODIAS had been presented on speech-related conferences in Germany, Great Britain, the Netherlands, Austria, and Italy. Various cooperation projects could be established , e.g. to the **University of Heidelberg** (in a deep brain electrical stimulation project with Parkinson's patients) or to the **University of Essen** (in a tumor surgery project).

Especially worth mentioning is the long-standing cooperation with the clinic for mouth-, jaw- and face-surgery of the Klinikum Rechts der Isar-Munich. The use of MODIAS resulted here in valuable experiences with patients suffering from organic (to be precise not neurogenic) speech disorders. In the framework of the world-wide most extensive mass screening of patients with cleft palates valuable information about the prevalence of voice disorders and about the effectiveness of new surgical techniques could be obtained.

Applications of MODIAS in the Clinical Neuropsychology Research Group (EKN) of the City Hospital **München-Bogenhausen** was directed at gaining basic information about sentence production and diadochokinesis in patients suffering from different neurogenic speech impairments. This included applications in long-term-studies on the treatment of patients with severe apraxia of speech. In this environment MODIAS has already become a **standard of clinical diagnostics and therapy**. In addition, it contributes, as a new research tool, to psycholinguistic, phonetic, and neuropsychological research.

Summary

The described clinical use has been the basis of several M.A. and doctoral theses, the results of which have been published in a number of scientific articles.

In the future, the clinical application of MODIAS shall be expanded. In addition to this, a migration to the latest MATLAB version is intended. A compilation of the MATLAB-based source code to a stand-alone-application running on any hardware is to follow, in order to further reduce system costs.

Literaturverzeichnis

[Ack89] H. Ackermann, W. Ziegler. *Die Dysarthroponie des Parkinson-Syndroms.* Fortschr. Neurol. Psychiat., 57, 149-160 (1989)

[Ack92] Ackermann, Ziegler. *Articulatory deficits in Parkinsonian dyarthria: an acoustic analysis.* Journal of Neurology, Neurosurgery and Psychiatrie, 54, 1093-8 (1992)

[Ahr92] T. Ahrndt, W. Ziegler. *Sprechen und Verstehen am Computer: Ein PC-basiertes Verfahren zur Verständlichkeitsprüfung mit dem Münchner Verständlichkeitsprofil (MVP).* Biomedical Journal 35, 4-8 (1992)

[Ahr94] T. Ahrndt. *Entwurf und Erprobung von Computersystemen zur Untersuchung zerebral bedingter Sprechstörungen.* Dissertation, eingereicht beim Institut für Nachrichtentechnik der Universität der Bundeswehr München (1994)

[Alf82] P.J. Alfonso, T. Baer. *Dynamics of vowel articulation.* LANG. SPEECH, 25, 151-173 (1982)

[Bak87] R.J. Baken. *Clinical measurement of speech and voice* College Hill Press, Boston (1987)

[Bas86] C. Bassich, C. Ludlow. *The use of perceptual methods by new clinicians for assessing voice quality.* Journal of Speech and Hearing Disorders, Vol. 51, pp. 125-133 (1986)

[Boe88] A. Böttiger. *Regelungstechnik.* Oldenbourg-Verlag, München (1988)

[Bor99] Jürgen Bortz. *Statistik für Sozialwissenschaftler.* Springer Verlag, 5. Auflage (1999)

[Bre97] T. Bressmann, M. Merk, R. Sader, W. Ziegler, H.H. Horch. *Computergestützte akustische Analyse bei Patienten mit Lippen-Kiefer-Gaumenspalten.* Biomedizinische Technik (Suppl. 2) 42 (1997)

[Bre98] T. Bressmann, R. Sader, M. Merk, W. Ziegler, R. Busch, H.-F. Zeilhofer, H.H. Horch. *Perzeptive und apparative Untersuchung an Patienten mit Lippen-Kiefer-Gaumenspalten.* Laryngo-Rhino-Otologie, Thieme-Verlag, Heft 12, Seite 700-708 (1998)

[Bro79] Bronstein, Semendjajew. *Taschenbuch der Mathematik.* Verlag Harri Deutsch (1979)

[Chi91] D.G. Childers, C.K. Lee. *Vocal quality factors: Analysis, synthesis and perception.* Journal of the Acoustical Society of America, 90, 2394-2410 (1991)

[Cry95] David Crystal. *Die Cambridge Enzyklopädie der Sprache.* Campus Verlag, Studienausgabe (1995)

[Dar75] F.L. Darley, A.E. Aronson and J.R. Brown. *Motor Speech Disorders.* W.B. Saunders, Philadelphia (1975)

[Dar85] J.K. Darby. *Epidemiology of neurologic diseases that produce communication disorders.* In J.K. Darby (Ed.), Speech and Language Evaluation in Neurology: Adult disorders. Grune and Stratton, Orlando, 29-44 (1985)

[Epp93] B. Eppinger, E. Herter. *Sprachverarbeitung.* Hanser-Verlag, Wien (1993)

[Esk90] L. Eskenazi, D.G. Childers, D.M. Hicks. *Acoustic correlates of vocal voice quality.* Journal of Speech, Language, and Hearing Research, Vol. 33, pp. 298-306 (1990)

[Fel84] K. Fellbaum. *Sprachverarbeitung und Sprachübertragung.* Springer-Verlag, Berlin (1984)

[Fin90] M. Finsterwald. *Methoden zur Diagnose und biofeedbackgesteuerten Therapie sprechmotorischer Störungen.* Dissertation, eingereicht beim Institut für Nachrichtentechnik der Universität der Bundeswehr München (1990)

[Fla58] J.L. Flanagan. *Some properties of the glottal sound source.* Journal of Speech, Language, and Hearing Research, Vol. 1, pp. 99-111 (1958)

[Gey84] H.-J. GEYWITZ. *Automatische Erkennung fließender Sprache mit silbenorientierten Einheiten.* Dissertation, Technische Universität München, (1984)

[Gilc79] E. Gilchrist, M. Wilkinson. *Some factors determining prognosis in young people with severe head injuries.* Arch. Neurol., 36, 355-359 (1979)

[Gilm81] S. Gilman, J.R. Bloedel, R. Lechtenberg. *Disorders of the Zerebellum.* Davis, Philadelphia (1981)

[Hes83] W. Hess. *Pitch Determination of Speech Signals.* Springer-Verlag (1983)

[IEEE79] IEEE. *Programms for Digital Signal Processing.* IEEE Press, Chapter 8 (1979)

[Hel1863] H.L.F. von Helmholtz. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik.* Verlag F. Vieweg und Sohn, Braunschweig (1863)

[Her99] J. Hermsdörfer, C. Marquardt, S. Wack, N. Mai. *Comparative analysis of diadochokinetic movements.* Journal of Electromyography and Kinesiology, 9, 283-295 (1999)

[Hil96] J. Hillenbrand, R. Houde. *Acoustic Correlates of Breathy Vocal Quality, Dysphonic Voices and Continuous Speech.* Journal of Speech, Language, and Hearing Research, Vol. 39, pp. 311-321 (1996)

[Hul98] W. Hulstijn, P.H.H.M. van Lieshout. *A motor skill approach to stuttering.* In W. Ziegler und K. Deger, Clinical Phonetics and Linguistics, London (1998)

[Hun80] M.J. Hunt, M. Lennig, P. Mermelstein. *Experiments in syllable-based recognition of continuous speech.* IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Denver, 880-883, (1980)

[Hus62] R. Husson. *Physiologie de la phonation.* Masson, Paris (1962)

[Hus73] H. Husquinet, M.C. Mackenzie-van der Noorda, N.C. Myrianthopoulos, H. Petit, W. Volkers, L.N. Went. *Analysis of Huntington's Chorea in Northwestern Europe.* Adv. Neurol., 1, 161-170 (1973)

[Kra94] T. Krauss, L. Shure, J. Little. *Signal Processing Toolbox for Use with MATLAB.* Math-Works Inc., Mass. (1994)

[Lea80] Lea, W.A. *Trends in speech recognition.* Prentice Hall, Englewood Cliffs, N.J., (1980)

[Lie77] P. Liebermann. *Speech physiology and acoustic phonetics.* Macmillan Publishing, New York (1977)

[Mak72] J. Makhoul, J. Wolf. *Linear prediction and the spectral analysis of speech.* Bolt Beranek and Newman, Cambridge, Mass., Rep. 2304 (1972)

[Mark72] J.D. Markel. *Digital Inverse Filtering- A new tool for formant trajectory estimation.* IEEE Trans. Audio Electroacoustic, vol. AU-20, pp. 129-137 (1972)

[Mark73] J.D. Markel. *Spectral Analysis of Speech by Linear Prediction.* IEEE Trans. Audio Electroacoustic, vol. AU-21, NO.3, pp. 140-148 (1973)

[Mars89] C.D. Marsden. *Slowness of movement in Parkinson's disease.* Movement Disorders 4. 26-37 (1989)

[Merm75] P. Mermelstein. *Automatic segmentation of speech into syllabic units.* JASA 58, 880-883 (1975)

[Merk97a] M. Merk, W. Ziegler, T. Schöbinger, K. Deger. *PC-gestützte Analyse phonetischer und phonematischer Aspekte der Sprachproduktion bei Aphasie und Sprechapraxie.* Jahrestagung der Arbeitsgemeinschaft für Aphasieforschung und -behandlung, Innsbruck, (1997)

[Merk97b] M. Merk, W. Ziegler. *A PC-based system for routine acoustic analysis of neurogenic speech disorders.* Poster präsentiert auf der 6.th annual conference of the International Clinical Phonetics and Linguistics Association ICPLA, Nijmegen, (13-15.10. 1997)

[Merk99a] M. Merk, W. Ziegler. *MODIAS - a PC-based System for routine analysis of neurogenic speech disorders.* In B. Massen and P. Groenen (Eds.), Pathologies of Speech and Language, Whurr Publishers, London (1999)

[Merk99b] M. Merk, W.Ziegler. *Acoustic assessment of neurogenic voice disorders in a clinical setting.* Vortrag, gehalten auf dem International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications, Florenz (1999)

[Mil87] P. Milenkovic. *Least mean square measures of voice perturbation.* Journal of Speech, Language, and Hearing Research, Vol. 30, pp. 529-538 (1987)

[Mue1840] J. Müller. *Handbuch der Physiologie des Menschen.* Verlag J. Hölscher, Koblenz (1840)

[Mum90] M. Mumenthaler. *Neurologie.* Thieme-Verlag (1990)

[Nil78] H.G. Nilsson. *A comparison of models for sharpening of the frequency selectivity in the coclea.* Biol. Cybernetics, 28, 177-181, (1978)

[Nor93] Marija J. Norusis. *SPSS for Windows, Professional Statistics.* SPSS Inc., Release 6.0 (1993)

[Not94] N.C. Notermans, G.W. Dijk, Y. Graaf, J. Gijn, J.H.J. Wokke. *Measuring ataxia: quantification based on the standard neurological examination.* Journal of Neurology, Neurosurgery, and Psychiatry, 57, 22-26. (1994)

[Ode90] K. Odell, M. Mc Neil, J.C. Rosenbeck, L. Hunter. *Perceptual characteristics of consonant production by apraxic speakers.* Journal of Speech and Hearing Disorders, 55 345-359 (1990)

[Par99] V. Parsa, D. Jamieson. *A Comparison of High Precision FO Extraction Algorithms for Sustained Vowels.* Journal of Speech, Language, and Hearing Research, Vol. 42, pp. 112-126 (1999)

[Per86] W. Perkins, R. Kent. *Textbook of functional anatomy of speech, language and hearing.* Taylor and Francis, London, Philadelphia (1986)

[Poe89] Klaus Poeck. *Klinische Neuropsychologie.* Thieme Verlag, Stuttgart (1989)

[Por82] R.A. Portnoy, A.E. Aronson. *Diadochokinetic syllable rate.* Journal of Speech and Hearing Disorders, 47, 324-328. (1982)

[Pro88] M. Prosiegel. *Beschreibung der Patientenstichprobe einer neuropsychologischen Rehabilitationsklinik.* In D. von Cramon, J. Zihl (Eds.) Neuropsychologische Rehabilitätionsverfahren, Springer-Verlag, Berlin (1988)

[Rab75] L.R. Rabiner, M.R. Sambur, C.E. Schmidt. *Applications of a Nonlinear Smoothing Algorithm to Speech Processing.* IEEE Trans. Acoust., Speech, and Signal Proc., vol. ASSP-23, No. 6, pp. 552-557 (1975)

[Rab78] N. Rabiner, W. Schafer. *Digital processing of speech signals.* Prentice Hall, (1978)

[Ram86] L.A. Ramig. *Acoustic analyses of phonation in patients with Huntington's disease.* Annals of Otology, Rhinology, and Laryngology 95. 288-293 (1986)

[Rie81] A.C.M. Rietveld. *A simples syllable detector.* Proceedings of the IRE, Vol. 48, No. 3 (1981)

[Rom98] Juliana Romberg. *Die Stimme.* GEO, Vol. 12, pp. 48-66 (1998)

[Rus92] Günther Ruske, Manfred Beham. *Gehörbezogene automatische Spracherkennung.* Erschienen in Mangold, Sprachliche Mensch-Maschine-Kommunikation, Oldenbourg Verlag, 33-47 (1992)

[Rus94] Günther Ruske. *Automatische Spracherkennung.* R. Oldenbourg Verlag, 2. Auflage (1994)

[Sar86] M.T. Sarno, A. Buonaguro, E. Levita. *Characteristics of verbal impairment in closed head injured patients.* Arch. Phys. Med. Rehabil., 67, 400-405 (1986)

[Sch95] R. Scherer, V. Vail, C. Geng Guo. *Required numbers of Tokens to Determine Representative Voice Perturbation Values.* Journal of Speech, Language, and Hearing Research, Vol. 38, pp. 1260-1269 (1995)

[Tei95] J. Teiwes. *PhonX, ein computerunterstütztes Experimentiersystem zur Analyse, Diagnostik und Therapie zentral bedingter Störungen von Sprachverarbeitungsprozessen.* Dissertation, eingereicht beim Institut für Nachrichtentechnik der Universität der Bundeswehr München (1995)

[Ter72] E. Terhardt. *Zur Tonhöhenwahrnehmung von Klängen.* ACUSTICA, Vol. 26 Heft 4, Verlag S. Hirzel, Stuttgart (1972)

[Tra88] H. Traunmüller. *Paralinguistic variation and invariance in the characteristic frequencies of vowels.* Phonetica, 45, 1-29 (1988)

[Van57] J. van den Berg, J.T. Zantema, P. Doornenbal. *On the air resistance and the Bernoulli-effect of the human larynx.* J. Acoust.Soc.Am. 29, 626-631 (1957)

[Vog75] A. Vogel. *Ein gemeinsames Funktionsschema zur Beschreibung der Lautheit und der Rauhigkeit.* Biol. Cybernetics 18, 31-40 (1975)

[Vog87] M. Vogel. *Einführung in die phonetische Beschreibung der Dysarthrophonien.* In L. Springer und G. Kattenbeck (Eds.), Aktuelle Beiträge zur Dysarthrophonie und Dysprosodie, tuduv Verlagsgesellschaft mbH, München (1987)

[Vog88] M. Vogel, W. Ziegler, H. Morasch. *Sprechen.* In D. Cramon und J. Zihl (Eds.), Neuropsychologische Rehabilitation, Springer-Verlag, Berlin, Heidelberg (1988)

[Yor84] K.M. Yorkston, D.R. Beukelman. *Assessment of Intelligibility of Dysarthric Speech.* C.C. Publications, Tigard (1984)

[Zie83a] W. Ziegler, D. von Cramon. *Vowel distortion in traumatic dysarthria: a formant study.* Phonetica, 40, 63-78 (1983)

[Zie83b] W. Ziegler, D. von Cramon. *Vowel distortion in traumatic dysarthria: Lip rounding versus tongue advancement.* Phonetica, 40, 312-322 (1983)

[Zie85] W. Ziegler, E. Hartmann, Ph. Hoole, D. von Cramon. *Entwicklung von diagnostischen Standards und von Therapieleitlinien für zentrale Stimm- und Sprechstörungen (Dysarthrophonien).* München: GSF (1985)

[Zie86] W. Ziegler, C. von Cramon. *Timing deficits in apraxia of speech.* European Archives of Psychiatry and Neurologic Sciences, 236, 44-49 (1986)

[Zie87] W. Ziegler, D. von Cramon. *Zentrale Stimmstörungen.* In L. Springer, G. Kattenbeck (Eds.) Interdisziplinäre Reihe zur Theorie und Praxis der Logopädie, Bd. 5, Tuduv-Verlag, München (1987)

[Zie88] Ziegler, Hoole, Hartmann, von Cramon. *Accelerated speech in dysarthria after acquired brain injury: Acoustic correlates.* British Journal of Disorders of Communication, 23, 215-28 (1988)

[Zie89] W. Ziegler, Ph. Hoole. *A combined acoustic and perceptual analysis of the tense-lax opposition in aphasic vowel production.* Aphasiology, 3, 449-463 (1989)

[Zie91] W. Ziegler. *Sprechapraktische Störungen bei Aphasie.* In G. Blanken (Ed.), Einführung in die linguistische Aphasiologie. (pp. 89-119). Freiburg: HochschulVerlag (1991)

[Zie93a] W. Ziegler, E. Hartmann. *Das Münchner Verständlichkeitsprofil (MVP); Untersuchungen zur Reliabilität und Validität.* Der Nervenarzt 64, 653-658 (1993)

[Zie93b] W. Ziegler, M. Vogel, H. Schröter-Morasch. *Sprechen.* In v. Cramon D., N. Mai, W. Ziegler (Eds), Neuropsychologische Diagnostik. Weinheim: VCH Verlag (1993)

[Zie93c] Ziegler, Hartmann, Hoole. *Syllabic timing in dysarthria.* Journal of Speech and Hearing Research, 36, 683-93 (1993)

[Zie94] W. Ziegler. *Prüfung der Verständlichkeit dysarthrischer Patienten: I. Grundlagen.* Sprache-Stimme-Gehör 18, 111-116 (1994)

[Zie96] W. Ziegler, K. Wessel. *Speech timing in ataxic disorders. Sentence production and rapid repetitive articulation.* Neurology, 47, 208-214. (1996)

[Zie98a] W. Ziegler, Mathias Vogel. *Diagnostik dysarthrischer Störungen.* In L. Springer und D. Schrey-Dern (Eds.), Dysarthrie (Grundlagen-Diagnostik-Therapie), Thieme-Verlag (1998)

[Zie98b] W. Ziegler. *Grundlagen der Dysarthrien.* In L. Springer und D. Schrey-Dern (Eds.), Dysarthrie (Grundlagen-Diagnostik-Therapie), Thieme-Verlag (1998)

[Zie98c] W. Ziegler. *Zentrale Kontrolle der Sprechmotorik.* Forum Lögopädie I 5-9 (1998)

[Zie99] W. Ziegler, P. Hoole. *Neurologic disease.* In R.D. Kent and M. Ball (Eds.), Handbook of Voice Quality Measurement, Singular Publishing Group, San Diego (1999)

[Zwi61] E. Zwicker. *Subdivision of the audible frequency range into critical bands (Frequenzgruppen).* JASA 33, 248 (1961)

[Zwi67] E. Zwicker, R. Feldtkeller. *Ohr als Nachrichtenempfänger.* Verlag S. Hirzel, Stuttgart (1967)

[Zwi77] E. Zwicker. *Procedure for calculating loudness of temporally variable sounds.* JASA, Vol. 62, 675-682 (1977)

[Zwi82] E. Zwicker. *Psychoakustik.* Verlag Springer, Berlin (1982)